

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2005, 11:51:56 ; Search time 165 seconds  
(without alignments)  
525.057 Million cell updates/sec

Title: US-10-623-429-9\_COPY\_404\_627

Perfect score: 1237  
Sequence: 1 AIAADROAGGLPAAAGDHGI.....ALVNASSAAHVNVDTARAAD 224

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_16Dec04.\*

- 1: geneseqp1980s.\*
- 2: geneseqp1990s.\*
- 3: geneseqp2000s.\*
- 4: geneseqp2001s.\*
- 5: geneseqp2002s.\*
- 6: geneseqp2003as.\*
- 7: geneseqp2003bs.\*
- 8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1237	100.0	329	6	Aae36429 HSV-2 UL2
2	1237	100.0	637	6	Aae36428 HSV-2 UL2
3	1237	100.0	637	8	Adj77800 Herpes si
4	1233	99.7	331	2	Aar64769 HSV-2 ISP
5	1233	99.7	636	2	Aar64767 HSV-2 pro
6	1233	99.7	638	2	Aar71017 HSV-2 pro
7	1233	99.7	642	2	Aar72124 HSV-2 str
8	1228	99.3	252	2	Aaw72049 HSV-2 str
9	1189	96.1	657	2	Aar72192 HSV-2 str
10	670.5	54.2	635	7	Aar71031 HSV-1 UL2
11	670.5	54.2	635	7	Adl18166 Human her
12	666.5	53.9	415	2	Aar28638 UL26 prot
13	666.5	53.9	603	2	Aar28641 UL26 prot
14	666.5	53.9	626	2	Aar28640 UL26 prot
15	666.5	53.9	635	2	Aar28634 UL26 prot
16	666.5	53.9	635	2	Aar28649 UL26 prot
17	666.5	53.9	635	2	Aar28647 UL26 prot
18	666.5	53.9	635	2	Aar28651 UL26 prot
19	666.5	53.9	635	2	Aar28648 UL26 prot
20	666.5	53.9	635	2	Aar28650 UL26 prot
21	666.5	53.9	635	2	Aar28652 UL26 prot
22	666.5	53.9	636	2	Aar28635 UL26 prot
23	661	53.4	636	2	Aar28636 UL26 prot
24	656	53.0	636	2	Aar28637 UL26 prot
25	623.5	50.4	615	2	Aar28642 UL26 prot

26	603.5	48.8	362	2	AAW69717	Aaw69717 Herpesvir
27	484	39.1	251	6	AAE36427	Aae36427 HSV-2 ant
28	399.5	32.3	514	2	AAR28643	Aar28643 UL26 prot
29	213.5	17.3	639	8	ADN21309	Adn21309 Bacterial
30	202	16.3	325	4	ABG21919	Abg21919 Novel hum
31	199	16.1	446	4	ABB70063	Abb70063 Drosophil
32	199	16.1	580	8	ABM83956	Abm83956 Human dia
33	199	16.1	580	8	ABM83957	Abm83957 Human dia
34	199	16.1	580	8	ABM83958	Abm83958 Human dia
35	197	15.9	205	6	ABU20105	Abu20105 Protein e
36	197	15.9	255	7	ADJ11524	Adj11524 Rice prot
37	196.5	15.9	1151	4	ABB61598	Abb61598 Drosophil
38	195	15.8	523	7	ADC81777	Adc81777 Human GPC
39	194.5	15.7	825	7	ADJ70479	Adj70479 Human hea
40	194.5	15.7	900	3	AAB42321	Aab42321 Human ORF
41	194.5	15.7	1217	4	ABG09876	Abg09876 Novel hum
42	194.5	15.7	1239	4	ABG09877	Abg09877 Novel hum
43	194.5	15.7	1638	7	ADC31236	Adc31236 Human nov
44	191	15.4	161	6	ABU19480	Abu19480 Protein e
45	191	15.4	406	4	ABG27250	Abg27250 Novel hum

## ALIGNMENTS

### RESULT 1

AAE36429  
ID AAE36429 standard; protein; 329 AA.

AC AAE36429;

XX 07-AUG-2003 (first entry)

XX HSV-2 UL26.5 full-length antigen.

XX Herpes simplex virus; HSV; infection; vaccine; therapy; UL26.5; antigen.

XX Herpes simplex virus type 2.

XX WO2003020108-A2.

XX 13-MAR-2003.

XX 27-AUG-2002; 2002WO-US027341.

XX 04-SEP-2001; 2001US-0317159P.

XX 17-APR-2002; 2002US-0373429P.

XX (CORI-) CORIXA CORP.

XX Hosken NA, Day CH;

XX WPI: 2003-290135/28.

XX N-PSDB; AAD55178.

XX Novel isolated polypeptide comprising immunogenic portion of a herpes simplex virus antigen, useful for detecting herpes simplex virus infection in a subject, and for treating the virus infection in a patient.

XX Claim 1; Page 112; 114pp; English.

XX The invention relates to polypeptides comprising an immunogenic portion of herpes simplex virus (HSV) antigen and to nucleic acid molecules encoding polypeptides. Polypeptides of the invention are useful for detecting and treating HSV infection in a patient. Polynucleotides of the invention are useful as diagnostic reagents for detecting HSV infection in a patient and also as probes or primers. The invention is used to prepare vaccines. The present sequence is HSV-2 strain HG52 UL26.5 full-length antigen

XX Sequence 329 AA;

XX SQ



Db 404 ATAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDRDFPFYYPGEARPEPRP 463  
QY 61 VDSRAARQASGPHETITLVCVAVTSLOQELAHMRARTHAPYGPYPVPVGYHHPHADTET 120  
Db 464 VDSRAARQASGPHETITLVCVAVTSLOQELAHMRARTHAPYGPYPVPVGYHHPHADTET 523  
QY 121 PAQPPRYPAKAVYLPPLPHIAPPGLPSGAVPPSPYPVAVTGPAPPLHQSPSAHAHPPP 180  
Db 524 PAQPPRYPAKAVYLPPLPHIAPPGLPSGAVPPSPYPVAVTGPAPPLHQSPSAHAHPPP 583  
QY 181 PPPGTPPPAASLPQEPAPGAEGALVNASSAAHVNDTARAAD 224  
Db 584 PPPGTPPPAASLPQEPAPGAEGALVNASSAAHVNDTARAAD 627

## RESULT 4

AAR64769  
ID AAR64769 standard; protein; 331 AA.

XX AC AAR64769;

XX XX 25-MAR-2003 (revised)  
DT 19-JUL-1995 (first entry)

XX DE HSV-2 ISP35 protein.

XX KW Protease; HSV-2; protease-inhibitor; virucide; ISP35.

XX OS Herpes simplex virus type 2.

XX PN WO9429456-A2.

XX XX 22-DEC-1994.

XX PF 25-MAY-1994; 94WO-US005920.

XX PR 08-JUN-1993; 93US-00073819.

XX PR 23-MAY-1994; 94US-00245390.

XX PA (ABBO ) ABBOTT LAB.

XX PI Steffy KR, Kati WM, Katz L, Mcgonigal TP, Sarthy AV, Schoen SE;

XX DR WPI; 1995-036483/05.

XX DR N-PSDB; AAQ76261.

XX XX New Herpes Simplex Virus type 2 protease - used in screening methods for  
PT identifying potential herpes viral protease inhibitor cpds.

XX PS Disclosure; Page 32-34; 56pp; English.

XX CC HSV-2 DNA (given in'AAQ76252) was isolated from Vero cells infected with  
CC HSV-2 strain Ci. The translated amino acid sequence is given in AAR64767.  
CC The DNA was digested with BamHI, sepd. by gel electrophoresis and  
CC transferred to nitrocellulose. A nick-translated probe deriv. from HSV-1  
CC protease was used for hybridization to identify DNA (AAQ76261) encoding  
CC HSV-2 protease (AAR64768) and the protease substrate ICP35 (AAR64769).  
CC (Updated on 25-MAR-2003 to correct PN field.)

XX XX Sequence 331 AA;

Query Match 99.7%; Score 1233; DB 2; Length 331;

Best Local Similarity 99.6%; Pred. No. 2.7e-84;

Matches 223; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDRDFPFYYPGEARPEPRP 60

Db 98 ATAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDRDFPFYYPGEARPEPRP 157

QY 61 VDSRAARQASGPHETITLVCVAVTSLOQELAHMRARTHAPYGPYPVPVGYHHPHADTET 120

Db 158 VDSRAARQASGPHETITLVCVAVTSLOQELAHMRARTHAPYGPYPVPVGYHHPHADTET 217

QY 121 PAQPPRYPAKAVYLPPLPHIAPPGLPSGAVPPSPYPVAVTGPAPPLHQSPSAHAHPPP 180  
Db 218 PAQPPRYPAEAVYLPPLPHIAPPGLPSGAVPPSPYPVAVTGPAPPLHQSPSAHAHPPP 277  
QY 181 PPPGTPPPAASLPQEPAPGAEGALVNASSAAHVNDTARAAD 224  
Db 278 PPPGTPPPAASLPQEPAPGAEGALVNASSAAHVNDTARAAD 321

## RESULT 5

AAR64767

ID AAR64767 standard; protein; 636 AA.

XX AC AAR64767;

XX XX 25-MAR-2003 (revised)  
DT 19-JUL-1995 (first entry)

XX XX HSV-2 protease, ICP35.

XX KW Protease; HSV-2; protease-inhibitor; virucide; ICP35.

XX OS Herpes simplex virus type 2.

XX PN WO9429456-A2.

XX PD 22-DEC-1994.

XX PF 25-MAY-1994; 94WO-US005920.

XX PR 08-JUN-1993; 93US-00073819.

XX PR 23-MAY-1994; 94US-00245390.

XX PA (ABBO ) ABBOTT LAB.

XX PI Steffy KR, Kati WM, Katz L, Mcgonigal TP, Sarthy AV, Schoen SE;

XX DR WPI; 1995-036483/05.

XX DR N-PSDB; AAQ76252.

XX XX New Herpes Simplex Virus type 2 protease - used in screening methods for  
PT identifying potential herpes viral protease inhibitor cpds.

XX PS Claim 4; Fig 1A-1C; 56pp; English.

XX CC HSV-2 DNA (given in AAQ76252) was isolated from Vero cells infected with  
CC HSV-2 strain Ci. The translated amino acid sequence is given in AAR64767.  
CC The DNA was digested with BamHI, sepd. by gel electrophoresis and  
CC transferred to nitrocellulose. A nick-translated probe deriv. from HSV-1  
CC protease was used for hybridization to identify DNA (AAQ76261) encoding  
CC HSV-2 protease (AAR64768) and the protease substrate ICP35 (AAR64769).  
CC (Updated on 25-MAR-2003 to correct PN field.)

XX XX Sequence 636 AA;

Query Match 99.7%; Score 1233; DB 2; Length 636;

Best Local Similarity 99.6%; Pred. No. 5.1e-84;

Matches 223; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDRDFPFYYPGEARPEPRP 60

Db 403 ATAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDRDFPFYYPGEARPEPRP 462

QY 61 VDSRAARQASGPHETITLVCVAVTSLOQELAHMRARTHAPYGPYPVPVGYHHPHADTET 120

Db 463 VDSRAARQASGPHETITLVCVAVTSLOQELAHMRARTHAPYGPYPVPVGYHHPHADTET 522

QY 121 PAQPPRYPAKAVYLPPLPHIAPPGLPSGAVPPSPYPVAVTGPAPPLHQSPSAHAHPPP 180

Db 523 PAQPPRYPAEAVYLPPLPHIAPPGLPSGAVPPSPYPVAVTGPAPPLHQSPSAHAHPPP 582

QY 181 PPPGTPPPAASLPQEPAPGAEGALVNASSAAHVNDTARAAD 224

Db 583 PPGPTPPAASLPQEPAPGAAGALVNASSAAHVNDTARAAD 626

## RESULT 6

AA71017  
ID AAR71017 standard; protein; 638 AA.

XX AC AAR71017;

XX DT 25-MAR-2003 (revised)

XX DT 02-OCT-1995 (first entry)

XX DE HSV-2 protease.

XX KW Herpes simplex virus; HSV-2; capsid; UL26 gene.

XX OS Herpes simplex virus type 2.

XX PN WO9506055-A1.

XX PD 02-MAR-1995.

XX PF 19-AUG-1994; 94WO-US009303.

XX PR 20-AUG-1993; 93US-00110522.

XX PR 23-JUN-1994; 94US-00264537.

XX PA (SMIK ) SMITHKLINE BEECHAM CORP.

XX PI Dillella AG, Debouck CW;

XX DR WPI; 1995-106803/14.

XX DR N-PSDB; AAQ84671.

XX PT New herpes simplex virus (HSV)-2 protease and capsid protein - used to develop prods. for use in the diagnosis and treatment of HSV-2 infections.

XX PS Claim 1; Fig 1; 51pp; English.

XX CC The sequence is the product of the herpes simplex virus type 2 gene UL26, which is the HSV-2 protease. The protein can be used in the diagnosis and treatment of HSV-2 infections. See also AAR71018-31. (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 638 AA;

Query Match 99.7%; Score 1233; DB 2; Length 638;  
Best Local Similarity 99.6%; Pred. No. 5.2e-84;  
Matches 223; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AIAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDRDPFYPYGEARPEPRP 60

Db 405 AIAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDRDPFYPYGEARPEPRP 464

Qy 61 VDSRRARQASGPHETITLVGAVTSLQELAHMRARTHAPYGPYPVGYHHPHADTET 120

Db 465 VDSRRARQASGPHETITLVGAVTSLQELAHMRARTHAPYGPYPVGYHHPHADTET 524

Qy 121 PAQPRYPKAVYLPYPPIAPPGLPGPLSGAVPPSPYVAVTPGAPPLHQPSPAHAHPPP 180

Db 525 PAQPRYPKAVYLPYPPIAPPGLPGPLSGAVPPSPYVAVTPGAPPLHQPSPAHAHPPP 584

Qy 181 PPGPTPPAASLPQEPAPGAAGALVNASSAAHVNDTARAAD 224

Db 585 PPGPTPPAASLPQEPAPGAAGALVNASSAAHVNDTARAAD 628

## RESULT 7

AA72124

ID AAW72124 standard; protein; 642 AA.

XX AC AAW72124;

XX DT 18-DEC-1998 (first entry)

XX DE HSV-2 strain SB5 Contig ID 15 ORF#11 protein.

XX KW HSV-2 strain SB5; immunological response induction; therapy; antiviral identification; viral protein inhibitor.

XX KW Herpes simplex virus 2.

XX OS WO9820016-A1.

XX PN 14-MAY-1998.

XX PF 31-OCT-1997; 97WO-US020016.

XX PR 04-NOV-1996; 96US-0030279P.

XX PR 09-JUN-1997; 97US-0049018P.

XX PA (SMIK ) SMITHKLINE BEECHAM CORP.

XX PI Esser XM, Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB; Leary JJ;

XX DR WPI; 1998-286847/25.

XX DR N-PSDB; AAV62159.

XX PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention and treatment of infection or inducing immunological response in mammal.

XX PS Claim 10; Page 92; 748pp; English.

XX CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein sequence of the invention. This sequence was isolated from a HSV-2 strain SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15. Based on homology, this sequence is a UL26 protein. The proteins can be used for the treatment or prevention of disease, to induce an immunological response in a mammal or to identify inhibitors, activators or novel antivirals. Antagonists of the proteins can be used to inhibit a viral polypeptide. The DNA sequence or a vector containing it can also be used to induce an immunological response in a mammal

XX SQ Sequence 642 AA;

Query Match 99.7%; Score 1233; DB 2; Length 642;  
Best Local Similarity 99.6%; Pred. No. 5.2e-84;  
Matches 223; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AIAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDRDPFYPYGEARPEPRP 60

Db 409 AIAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDRDPFYPYGEARPEPRP 468

Qy 61 VDSRRARQASGPHETITLVGAVTSLQELAHMRARTHAPYGPYPVGYHHPHADTET 120

Db 469 VDSRRARQASGPHETITLVGAVTSLQELAHMRARTHAPYGPYPVGYHHPHADTET 528

Qy 121 PAQPRYPKAVYLPYPPIAPPGLPGPLSGAVPPSPYVAVTPGAPPLHQPSPAHAHPPP 180

Db 529 PAQPRYPKAVYLPYPPIAPPGLPGPLSGAVPPSPYVAVTPGAPPLHQPSPAHAHPPP 588

Qy 181 PPGPTPPAASLPQEPAPGAAGALVNASSAAHVNDTARAAD 224

Db 589 PPGPTPPAASLPQEPAPGAAGALVNASSAAHVNDTARAAD 632

## RESULT 8

AA72049

ID AAW72049 standard; protein; 252 AA.

XX AC AAW72049;

XX DT 07-DEC-1998 (first entry)

DE HSV-2 strain SB5 Contig ID 104 ORF#17 protein.  
KW HSV-2 strain SB5; immunological response induction; therapy;  
XX antiviral identification; viral protein inhibitor.  
KW WO9820016-A1.  
XX Herpes simplex virus 2.  
XX PD 14-MAY-1998.  
XX 31-OCT-1997; 97WO-US020016.  
XX 04-NOV-1996; 96US-0030279P.  
XX 09-JUN-1997; 97US-0049018P.  
XX (SMIK ) SMITHKLINE BEECHAM CORP.  
XX Esser KM, Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB,  
XX Leary JJ;  
XX WPI: 1998-286847/25.  
XX N-PSDB; AAV62134.  
XX Herpes simplex virus type-2 sequences - useful in, e.g. prevention and  
XX treatment of infection or inducing immunological response in mammal.  
XX Claim 10; Page 58-59; 748pp; English.  
XX This sequence represents a Herpes simplex virus type-2 (HSV-2) protein  
XX sequence of the invention. This sequence was isolated from a HSV-2 strain  
XX SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 104.  
XX Based on homology, this sequence is a UL26 protease. The proteins can be  
XX used for the treatment or prevention of disease, to induce an  
XX immunological response in a mammal or to identify inhibitors, activators  
XX or novel antivirals. Antagonists of the proteins can be used to inhibit a  
XX viral polypeptide. The DNA sequence or a vector containing it can also be  
XX used to induce an immunological response in a mammal  
SQ Sequence 252 AA;  
Query Match 99.3%; Score 1228; DB 2; Length 252;  
Best Local Similarity 99.1%; Pred. No. 5e-84;  
Matches 222; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AIAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDRDFPYPGGEARPEPRP 60  
DB 19 AIAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDRDFPYPGGEARPEPRP 78  
QY 61 VDSRAARQAQSGPHETITLGVAVTSLQQLAHMRARTHAPYGPYPVPVGYHHPHADTET 120  
DB 79 VDSRAARQAQSGPHETITLGVAVTSLQQLAHMRARTHAPYGPYPVPVGYHHPHADTET 138  
QY 121 PAQPPRYPAKAVYLPPIHAPPGPLSGAVPPSPVPVAVTGPAPPLHQPSPAHAPPP 180  
DB 139 PAQPPRYPAEAVYLPPIHAPPGPLSGAVPPSPVPVAVTGPAPPLHQPSPAHAPPP 198  
QY 181 PPPGTPPPPAASLPQEPAGAGALVNASSAAHV 224  
DB 199 PPPGTPPPPAASLPQEPAGAGALVNASSAAHV 242  
RESULT 9  
AAW72192  
ID AAW72192 standard; protein; 657 AA.  
XX AAW72192;  
XX 13-JAN-1999 (first entry)  
XX HSV-2 strain SB5 Contig ID 15 ORF#27 protein.  
XX HSV-2 strain SB5; immunological response induction; therapy;  
KW

KW antiviral identification; viral protein inhibitor.  
XX Herpes simplex virus 2.  
XX WO9820016-A1.  
XX 14-MAY-1998.  
XX 31-OCT-1997; 97WO-US020016.  
XX 04-NOV-1996; 96US-0030279P.  
XX 09-JUN-1997; 97US-0049018P.  
XX (SMIK ) SMITHKLINE BEECHAM CORP.  
XX Esser KM, Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB,  
XX Leary JJ;  
XX WPI: 1998-286847/25.  
XX N-PSDB; AAV62176.  
XX Herpes simplex virus type-2 sequences - useful in, e.g. prevention and  
XX treatment of infection or inducing immunological response in mammal.  
XX Claim 10; Page 120; 748pp; English.  
XX This sequence represents a Herpes simplex virus type-2 (HSV-2) protein  
XX sequence of the invention. This sequence was isolated from a HSV-2 strain  
XX SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.  
XX Based on homology, this sequence is a capsid protein p40. The proteins  
XX can be used for the treatment or prevention of disease, to induce an  
XX immunological response in a mammal or to identify inhibitors, activators  
XX or novel antivirals. Antagonists of the proteins can be used to inhibit a  
XX viral polypeptide. The DNA sequence or a vector containing it can also be  
XX used to induce an immunological response in a mammal  
SQ Sequence 657 AA;  
Query Match 96.1%; Score 1189; DB 2; Length 657;  
Best Local Similarity 99.5%; Pred. No. 1e-80;  
Matches 214; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AIAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDRDFPYPGGEARPEPRP 60  
DB 405 AIAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDRDFPYPGGEARPEPRP 464  
QY 61 VDSRAARQAQSGPHETITLGVAVTSLQQLAHMRARTHAPYGPYPVPVGYHHPHADTET 120  
DB 465 VDSRAARQAQSGPHETITLGVAVTSLQQLAHMRARTHAPYGPYPVPVGYHHPHADTET 524  
QY 121 PAQPPRYPAKAVYLPPIHAPPGPLSGAVPPSPVPVAVTGPAPPLHQPSPAHAPPP 180  
DB 525 PAQPPRYPAEAVYLPPIHAPPGPLSGAVPPSPVPVAVTGPAPPLHQPSPAHAPPP 584  
QY 181 PPPGTPPPPAASLPQEPAGAGALVNASSAAHV 215  
DB 585 PPPGTPPPPAASLPQEPAGAGALVNASSAAHV 619  
RESULT 10  
AAW71031  
ID AAW71031 standard; protein; 635 AA.  
XX AAW71031;  
XX 25-MAR-2003 (revised)  
XX 02-OCT-1995 (first entry)  
XX HSV-1 UL26 gene product.  
XX Herpes simplex virus; HSV-1 protease; capsid.  
XX Herpes simplex virus type 1.

```

XX PN WO9506055-A1.
XX PD 02-MAR-1995.
XX PR 19-AUG-1994; 94WO-US009303.
XX PR 20-AUG-1993; 93US-00110522.
XX PR 23-JUN-1994; 94US-00264537.
XX PA (SMIK ) SMITHLINE BEECHAM CORP.
XX PI Dillella AG, Debouck CM;
XX DR WPI; 1995-106803/14.
XX DR N-PSDB; AAQ84678.
XX PT New herpes simplex virus (HSV)-2 protease and capsid protein - used to
XX PT develop prods. for use in the diagnosis and treatment of HSV-2
XX PT infections.
XX PS Disclosure; Page 35; 51pp; English.
XX SS The sequence is that of a herpes simplex virus type 1 protease and
XX CC capsid. The protein can be used in the diagnosis and treatment of HSV-1
XX CC infections. See also R717017-30. (Updated on 25-MAR-2003 to correct FN
XX CC field.)
XX SQ Sequence 635 AA;
Query Match 54.2%; Score 670.5; DB 2; Length 635;
Best Local Similarity 63.4%; Pred. No. 5.7e-42;
Matches 144; Conservative 12; Mismatches 64; Indels 7; Gaps 4;
QY 1 AIAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDPRDPFYYPGEARPERP 60
DB 403 AIAADROAGGQP--AAGDPGVRGSKRRRYEAGPSESVCDDQDEPDADYPYYPGEARGAPRG 461
QY 61 VDSRRARQASGPHETITLAVGAVTSIQQLAHMRARTHAPYGPYPVGYPHPHADTE- 119
DB 62 VDSRRARHSPGTNETITLALMGAVTSIQQLAHMRARTSAPYGMYPVAHYRQVQSGPEP 521
QY 120 TPAQPPRYPAKAVYLPPIPHIAPGPPLSGA--VPPRSYPPVAVTPGAPPLHOPSPAH 177
DB 522 TTHPALCPPEAVTRPPHAPYGPPOGPASHAPTPPYAPACPPGPPP---PCPSTQT 578
QY 178 PPPPPPGFTPPPAASLPQPEAPGAEAGALYNASSAAHVNDTARAAD 224
DB 579 RAPLPTPEAFPFAATGSPQPEASNAEAGALYNASSAAHVNDTARAAD 625
RESULT 11
ADL18166
XX ID ADL18166 standard; protein; 635 AA.
XX AC ADL18166;
XX XX
XX XX 06-MAY-2004 (first entry)
XX XX Human herpesvirus 1 protease protein SEQ ID NO:86.
XX DE
XX KW chimeric protein; signal protein; trafficking signal targeting;
XX KW proteolytic cleavage site; protease; protease inhibitor; enzyme.
XX OS Human herpesvirus 1.
XX XX
XX PN WO2003014381-A1.
XX PD 20-FEB-2003.
XX PF 08-AUG-2002; 2002WO-KR001515.
XX PR 10-AUG-2001; 2001KR-00048123.
XX
XX (AHRA-) AHRA BIOSYSTEMS INC.
XX PI Hwang I, Kim DH, Lee YJ;
XX DR WPI; 2003-256596/25.
XX DR N-PSDB; ADL18165.
XX PT New chimeric protein, useful for detecting protease inhibitors inside the
XX PT cell or tissue.
XX PS Disclosure; SEQ ID NO 86; 214pp; English.
XX SS The present invention describes a chimeric protein comprising at least
XX CC one signal protein that has a trafficking signal targeting to a
XX CC subcellular organelle and at least one proteolytic cleavage site for a
XX CC protease. The chimeric protein is constructed, so that: (a) the
XX CC trafficking signals of all the signal proteins are inactivated by linking
XX CC the proteolytic site or a signal masking protein through the proteolytic
XX CC site to the N-or C- terminus of the signal proteins, and so the chimeric
XX CC protein is present in cytosol; (b) the trafficking signal of at least one
XX CC signal protein is activated when the proteolytic cleavage site is cleaved
XX CC by the protease, and as a result at least one fragment protein that
XX CC includes the activated signal protein is a transported to a subcellular
XX CC organelle; and (c) the chimeric protein is labelled with at least one
XX CC fluorescent protein and the position and intensity distribution of the
XX CC fluorescent label signal in the cell is altered depending on the cleavage
XX CC by the protease. Also described: (1) a recombinant gene comprising a
XX CC nucleic acid sequence encoding the chimeric protein which is constructed
XX CC to express the chimeric protein in a cell; (2) a cell transformed with
XX CC the recombinant gene or vector; (3) analysing the activity of a protease
XX CC in vivo; (4) screening protease inhibitors in vivo; (5) a system for
XX CC detecting a protease inside a cell; (6) a nucleic acid comprising the
XX CC sequence encoding the chimeric protein for detecting protease activity in
XX CC a cell; (7) a vector comprising the nucleic acid; (8) a kit for detecting
XX CC a protease inside a cell comprising the chimeric protein or the vector;
XX CC (9) detecting a protease inside a cell or tissue; and (10) detecting a
XX CC protease inhibitor in vivo. The chimeric protein is useful for detecting
XX CC protease inhibitors inside the cell or tissue. The present sequence
XX CC represents HSV-1 protease, which is used in the exemplification of the
XX CC present invention.
XX SQ Sequence 635 AA;
Query Match 54.2%; Score 670.5; DB 7; Length 635;
Best Local Similarity 63.4%; Pred. No. 5.7e-42;
Matches 144; Conservative 12; Mismatches 64; Indels 7; Gaps 4;
QY 1 AIAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPDPRDPFYYPGEARPERP 60
DB 403 AIAADROAGGQP--AAGDPGVRGSKRRRYEAGPSESVCDDQDEPDADYPYYPGEARGAPRG 461
QY 61 VDSRRARQASGPHETITLAVGAVTSIQQLAHMRARTHAPYGPYPVGYPHPHADTE- 119
DB 462 VDSRRARHSPGTNETITLALMGAVTSIQQLAHMRARTSAPYGMYPVAHYRQVQSGPEP 521
QY 120 TPAQPPRYPAKAVYLPPIPHIAPGPPLSGA--VPPRSYPPVAVTPGAPPLHOPSPAH 177
DB 522 TTHPALCPPEAVTRPPHAPYGPPOGPASHAPTPPYAPACPPGPPP---PCPSTQT 578
QY 178 PPPPPPGFTPPPAASLPQPEAPGAEAGALYNASSAAHVNDTARAAD 224
DB 579 RAPLPTPEAFPFAATGSPQPEASNAEAGALYNASSAAHVNDTARAAD 625
RESULT 12
AAR28638
XX ID AAR28638 standard; protein; 415 AA.
XX AC AAR28638;
XX XX
XX XX 25-MAR-2003 (revised)
XX DT 24-MAR-1993 (first entry)

```

XX UL26 protease deletion mutant D, amino acids 1-220 deleted.

XX UL26; open reading frame; ORF; protease; herpes simplex virus; HSV;

XX capsid protein; ICP35; ICP35 c; ICP35 d; ICP35 e; ICP35 f;

XX functional domain; insertion; deletion; substitution.

XX Herpes simplex virus type 1.

XX EP514830-A2.

XX 25-NOV-1992.

XX 19-MAY-1992; 92EP-00108420.

XX 24-MAY-1991; 91US-00705814.

XX 07-FEB-1992; 92US-00832855.

XX (ARCH-) ARCH DEV CORP.

XX Roizman B, Liu F;

XX WPI; 1992-391444/48.

XX New herpes serine proteases and corresp. nucleic acid sequences - for

XX detection, prevention and treatment of infection caused by HSV, Epstein-

XX Barr, Varicella-Zoster and CMV cytomegalo-virus.

XX Disclosure; Table 1; 66pp; English.

XX The sequences given in AAR28635-52 are mutants of the protease encoded by

XX the UL26 open reading frame (ORF). The UL26 ORF is contained within a

XX section of the herpes simplex virus (HSV) type 1 genome for the family of

XX HSV capsid proteins designated ICP35. The UL26 transcription initiation

XX site is used as the start point for measurements of distance in this

XX section of the genome. The UL26 protease and ICP35 precursor proteins,

XX ICP35 c and d, may be cleaved by the UL26 protease. Cleavage of ICP35 c

XX and d gives ICP35 e and f respectively. Cleavage specificity of the UL26

XX protease is altered by the inclusion of these mutations. These mutations

XX include insertions, deletions or substitutions which affect the functional

XX domains of the protease. (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 415 AA;

Query Match 53.9%; Score 666.5; DB 2; Length 415;

Best Local Similarity 63.0%; Pred. No. 7.5e-42; Mismatches 65; Indels 7; Gaps 4;

Matches 143; Conservative 12; Mismatches 65; Indels 7; Gaps 4;

QY 1 AIAADROAGGLPAAAGDHGIRGSAKRHRHEVEQPEYDCGRDEPDRDFPYYPGEARPEPRP 60

DB 183 AIAADROAGGQP-AAGDPGVRGSGKRRRYEAGPSESICYDQDEPDADYPYYPGEARGPRG 241

QY 61 VDSRRAARQASGPHTTITALVGAVTSLOQELAHMARTHAPYGPYPVPVGHHPHADTE- 119

DB 242 VDSRRAARHS PGTNETITATMGAVTSLOQELAHMARTSAPYGMTTPVAHYRPQVGEPEP 301

QY 120 TPAQPRPRPAKAVYLPPIHAPPGPLS--GAVPPSPYPPVAVTGPAPPLHQPSPAH 177

DB 302 TTTTHALCPPEAVYRPPPHSAPYGPQPPSHAPTPPYAPACPGPPPP---PCPSTQT 358

QY 178 PPPPPPGTTPPAAASLPQEPAGAGALVNASSAAHNVNDTARAAD 224

DB 359 RAPLPEPAFPFAATGSPQPEASNAEAGALVNASSAAHNVNDTARAAD 405

RESULT 13

AAR28641

ID AAR28641 standard; protein; 603 AA.

XX AAR28641;

XX 25-MAR-2003 (revised)

DT 24-MAR-1993 (first entry)

XX UL26 protease deletion mutant FF, amino acids 1-32 deleted.

XX UL26; open reading frame; ORF; protease; herpes simplex virus; HSV;

XX capsid protein; ICP35; ICP35 c; ICP35 d; ICP35 e; ICP35 f;

XX functional domain; insertion; deletion; substitution.

XX Herpes simplex virus type 1.

XX EP514830-A2.

XX 25-NOV-1992.

XX 19-MAY-1992; 92EP-00108420.

XX 24-MAY-1991; 91US-00705814.

XX 07-FEB-1992; 92US-00832855.

XX (ARCH-) ARCH DEV CORP.

XX Roizman B, Liu F;

XX WPI; 1992-391444/48.

XX New herpes serine proteases and corresp. nucleic acid sequences - for

XX detection, prevention and treatment of infection caused by HSV, Epstein-

XX Barr, Varicella-Zoster and CMV cytomegalo-virus.

XX Disclosure; Table 1; 66pp; English.

XX The sequences given in AAR28635-52 are mutants of the protease encoded by

XX the UL26 open reading frame (ORF). The UL26 ORF is contained within a

XX section of the herpes simplex virus (HSV) type 1 genome for the family of

XX HSV capsid proteins designated ICP35. The UL26 transcription initiation

XX site is used as the start point for measurements of distance in this

XX section of the genome. The UL26 protease and ICP35 precursor proteins,

XX ICP35 c and d, may be cleaved by the UL26 protease. Cleavage of ICP35 c

XX and d gives ICP35 e and f respectively. Cleavage specificity of the UL26

XX protease is altered by the inclusion of these mutations. These mutations

XX include insertions, deletions or substitutions which affect the functional

XX domains of the protease. (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 603 AA;

Query Match 53.9%; Score 666.5; DB 2; Length 603;

Best Local Similarity 63.0%; Pred. No. 1.1e-41; Mismatches 143; Conservative 12; Mismatches 65; Indels 7; Gaps 4;

QY 1 AIAADROAGGLPAAAGDHGIRGSAKRHRHEVEQPEYDCGRDEPDRDFPYYPGEARPEPRP 60

DB 371 AIAADROAGGQP-AAGDPGVRGSGKRRRYEAGPSESICYDQDEPDADYPYYPGEARGPRG 429

QY 61 VDSRRAARQASGPHTTITALVGAVTSLOQELAHMARTHAPYGPYPVPVGHHPHADTE- 119

DB 430 VDSRRAARHS PGTNETITATMGAVTSLOQELAHMARTSAPYGMTTPVAHYRPQVGEPEP 489

QY 120 TPAQPRPRPAKAVYLPPIHAPPGPLS--GAVPPSPYPPVAVTGPAPPLHQPSPAH 177

DB 490 TTTTHALCPPEAVYRPPPHSAPYGPQPPSHAPTPPYAPACPGPPPP---PCPSTQT 546

QY 178 PPPPPPGTTPPAAASLPQEPAGAGALVNASSAAHNVNDTARAAD 224

DB 547 RAPLPEPAFPFAATGSPQPEASNAEAGALVNASSAAHNVNDTARAAD 593

RESULT 14

AAR28640

ID AAR28640 standard; protein; 626 AA.

XX AAR28640;

XX 25-MAR-2003 (revised)

DT 24-MAR-1993 (first entry)

```

XX UL26 protease deletion mutant EE, amino acids 1-9 deleted.
DE
XX
XX UL26; open reading frame; ORF; protease; herpes simplex virus; HSV;
KW capsid protein; ICP35 c; ICP35 d; ICP35 e; ICP35 f;
KW functional domain; insertion; deletion; substitution.
OS
XX Herpes simplex virus type 1.
XX
XX EP514830-A2.
PN
XX
XX 25-NOV-1992.
PD
XX
XX 19-MAY-1992; 92EP-00108420.
PF
XX
XX 24-MAY-1991; 91US-00705814.
PR
XX 07-FEB-1992; 92US-00832855.
PR
XX
XX (ARCH-) ARCH DEV CORP.
PA
XX
XX Roizman B, Liu F;
PI
XX
XX WPI; 1992-391444/48.
DR
XX
XX New herpes serine proteases and corresp. nucleic acid sequences - for
PT detection, prevention and treatment of infection caused by HSV, Epstein-
PT Barr, Varicella-Zoster and CMV cytomegalo-virus.
XX
XX Disclosure; Table 1; 66pp; English.
PS
XX
XX The sequences given in AAR28635-52 are mutants of the protease encoded by
CC the UL26 open reading frame (ORF). The UL26 ORF is contained within a
CC section of the herpes simplex virus (HSV) type 1 genome for the family of
CC HSV capsid proteins designated ICP35. The UL26 transcription initiation
CC site is used as the start point for measurements of distance in this
CC section of the genome. The UL26 protease and ICP35 precursor proteins,
CC ICP35 c and d, may be cleaved by the UL26 protease. Cleavage of ICP35 c
CC and d gives ICP35 e and f respectively. Cleavage specificity of the UL26
CC protease is altered by the inclusion of these mutations. These mutations
CC include insertions, deletions or substitutions which affect the functional
CC domains of the protease. (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 626 AA;
SQ
Query Match 53.9%; Score 666.5; DB 2; Length 626;
Best Local Similarity 63.0%; Pred. No. 1.1e-41;
Matches 143; Conservative 12; Mismatches 65; Indels 7; Gaps 4;
QY 1 AIAADRQAGGLPAAAGDHGIRGSAKRRRHEVEQPEYDCGRDEPRDPFPYYPGGEARPEPRP 60
DB 394 AIAADRQAGQP-AAGDPGVRGSGKRRRYEAGSPSESYCDQDEPDADYPYYPGGEARGAPRG 452
QY 61 VDSRRARQASGPHETITLGVAVTSLQQLAHMRARTHAPYGPYPVPVGYHHPHADTE- 119
DB 453 VDSRRARHSFGNETITLGMGAVTSLQQLAHMRARTSAPYGMVTFVAHYRQVGEPEP 512
QY 120 TPAQPPRYPAKAVLYLPPPHIAPPGLPS--GAVPPPSYPPVAVTPGAPPLHQPSPAHAH 177
DB 513 TTHPALCPPEAVYRPPHSAFYGPQGPSPHAPTTPYAPACFPGPPPP--PCPSTQT 569
QY 178 PPPPPGPTPPPAASLPQPEAPCAEAGALVNASSAAHVNDTARAAD 224
DB 570 RAPLPTPEAPPAPTAATGSQPEASNAEAGALVNASSAAHVNDTARAAD 616
RESULT 15
AAR28634
ID AAR28634 standard; protein; 635 AA.
XX
XX AAR28634;
XX
XX 25-MAR-2003 (revised)
DT 24-MAR-1993 (first entry)

```

```

XX UL26 protease.
DE
XX
XX UL26; open reading frame; ORF; protease; herpes simplex virus; HSV;
KW capsid protein; ICP35; ICP35 c; ICP35 d; ICP35 e; ICP35 f;
KW functional domain.
OS
XX Herpes simplex virus type 1.
XX
XX EP514830-A2.
PN
XX
XX 25-NOV-1992.
PD
XX
XX 19-MAY-1992; 92EP-00108420.
PF
XX
XX 24-MAY-1991; 91US-00705814.
PR
XX 07-FEB-1992; 92US-00832855.
PR
XX
XX (ARCH-) ARCH DEV CORP.
PA
XX
XX Roizman B, Liu F;
PI
XX
XX WPI; 1992-391444/48.
DR
XX
XX N-PSDB; AAQ30736.
DR
XX
XX New herpes serine proteases and corresp. nucleic acid sequences - for
PT detection, prevention and treatment of infection caused by HSV, Epstein-
PT Barr, Varicella-Zoster and CMV cytomegalo-virus.
XX
XX Disclosure; Fig 1B; 66pp; English.
PS
XX
XX This sequence is encoded by the UL26 open reading frame (ORF) and it is a
CC protease. The UL26 ORF is contained within a section of the herpes
CC simplex virus (HSV) type 1 genome for the family of HSV capsid proteins
CC designated ICP35. The UL26 transcription initiation site is used as the
CC start point for measurements of distance in this section of the genome.
CC The UL26 protease and ICP35 precursor proteins, ICP35 c and d, may be
CC cleaved by the UL26 protease. Cleavage of ICP35 c and d gives ICP35 e and
CC f respectively. Cleavage specificity of the UL26 protease may be altered
CC by mutations within the sequence. These mutations may be insertions,
CC deletions or substitutions which affect the functional domains of the
CC protease. See also AAR28635-52. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
XX Sequence 635 AA;
SQ
Query Match 53.9%; Score 666.5; DB 2; Length 635;
Best Local Similarity 63.0%; Pred. No. 1.1e-41;
Matches 143; Conservative 12; Mismatches 65; Indels 7; Gaps 4;
QY 1 AIAADRQAGGLPAAAGDHGIRGSAKRRRHEVEQPEYDCGRDEPRDPFPYYPGGEARPEPRP 60
DB 403 AIAADRQAGQP-AAGDPGVRGSGKRRRYEAGSPSESYCDQDEPDADYPYYPGGEARGAPRG 461
QY 61 VDSRRARQASGPHETITLGVAVTSLQQLAHMRARTHAPYGPYPVPVGYHHPHADTE- 119
DB 462 VDSRRARHSFGNETITLGMGAVTSLQQLAHMRARTSAPYGMVTFVAHYRQVGEPEP 521
QY 120 TPAQPPRYPAKAVLYLPPPHIAPPGLPS--GAVPPPSYPPVAVTPGAPPLHQPSPAHAH 177
DB 522 TTHPALCPPEAVYRPPHSAFYGPQGPSPHAPTTPYAPACFPGPPPP--PCPSTQT 578
QY 178 PPPPPGPTPPPAASLPQPEAPCAEAGALVNASSAAHVNDTARAAD 224
DB 579 RAPLPTPEAPPAPTAATGSQPEASNAEAGALVNASSAAHVNDTARAAD 625

```

Search completed: July 27, 2005, 12:18:38

Job time : 167 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2005, 12:31:33 ; Search time 39 Seconds  
(without alignments)  
22.204 Million cell updates/sec

Title: US-10-623-429-9\_COPY\_475\_483  
Perfect score: 48  
Sequence: 1 GPHETITAL 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

PIR\_79:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	79.2	83	2 S24712	Ig alpha chain - h
2	37	77.1	404	2 T35421	probable regulator
3	36	75.0	459	2 E86726	cell division prot
4	36	75.0	1825	2 C88400	protein H19W22.1 l
5	36	75.0	1825	2 T32828	hypothetical prote
6	35	72.9	201	2 A12205	hypothetical prote
7	35	72.9	334	2 AB0936	N-acetyl-gamma-glu
8	35	72.9	515	1 T06595	2-amino-4-hydroxy-
9	35	72.9	724	2 B71404	hypothetical prote
10	35	72.9	1239	1 VHWVEE	structural polypro
11	35	72.9	1240	1 VHWVEE	structural polypro
12	35	72.9	1241	2 S26373	genome polyprotein
13	35	72.9	1242	2 A56605	structural polypro
14	35	72.9	1242	2 S72350	structural polypro
15	35	72.9	1324	2 T14070	peptide synthetase
16	34	70.8	143	2 A70573	hypothetical prote
17	34	70.8	478	2 T40860	probable alpha-amy
18	34	70.8	653	2 G82971	probable ferredoxi
19	33	68.8	181	2 D87196	probable membrane
20	33	68.8	259	2 C70530	hypothetical prote
21	33	68.8	384	2 AD0049	hypothetical prote
22	33	68.8	418	2 E69014	hypothetical prote
23	33	68.8	511	2 D69675	phosphoglycerate m
24	33	68.8	557	2 D86106	yidB protein [semi
25	33	68.8	557	2 S56342	yidB protein - Esc
26	33	68.8	557	2 H91265	hypothetical prote
27	33	68.8	613	2 D87380	hypothetical prote
28	33	68.8	633	2 S49788	probable membrane
29	33	68.8	635	1 WNBEM6	capsid protein - h

30	68.8	706	2	D82452	anaerobic ribonucel
31	68.8	707	1	A64047	ribonucleoside-tri
32	68.8	712	1	A47331	ribonucleoside-tri
33	68.8	712	2	AC1058	ribonucleoside-tri
34	68.8	712	2	G91280	anaerobic ribonucel
35	68.8	712	2	G86121	anaerobic ribonucel
36	68.8	712	2	AG0419	ribonucleoside-tri
37	68.8	971	2	S54595	probable membrane
38	68.8	1008	2	T30544	major surface glyc
39	68.8	1017	2	T30542	major surface glyc
40	68.8	1022	2	T30543	major surface glyc
41	68.8	1179	2	C36792	hypothetical prote
42	68.8	1267	1	MXR31	lambda 3 protein -
43	68.8	1267	1	MXR32	lambda 3 protein -
44	68.8	1267	1	MXR33	lambda 3 protein -
45	67.7	467	2	T38353	serine hydroxymeth
46	67.7	472	2	T37918	serine hydroxymeth
47	67.7	479	1	A42241	glycine hydroxymet
48	66.7	117	2	T03782	probable lipid tra
49	66.7	118	2	T02042	lipid transfer pro
50	66.7	118	2	T03300	probable lipid tra
51	66.7	141	2	H75583	probable copper re
52	66.7	168	2	B69199	hypothetical prote
53	66.7	179	2	C70507	hypothetical prote
54	66.7	343	2	C90435	hypothetical prote
55	66.7	404	2	E83851	N-acyl-L-amino aci
56	66.7	425	2	B98012	cell division prot
57	66.7	429	2	D95144	hypothetical prote
58	66.7	466	2	PC4296	nicotinic acetylch
59	66.7	479	2	D70641	hypothetical prote
60	66.7	554	2	T08976	2-amino-4-hydroxy
61	66.7	579	2	B44857	acetylactate synth
62	66.7	680	2	T03106	probable transport
63	66.7	680	2	T08080	probable myrosinas
64	66.7	858	2	E72419	flagellar hook-ass
65	66.7	912	2	T18785	hypothetical prote
66	66.7	1213	2	T37959	hypothetical prote
67	66.7	2035	2	AC0233	yersiniabactin bio
68	66.7	2035	2	A48654	probable polyketid
69	66.7	2041	2	T17439	peptide synthetase
70	66.7	29	2	S34762	L-serine ammonia-1
71	64.6	98	2	D95371	hypothetical prote
72	64.6	132	2	G82805	two-component syst
73	64.6	139	2	S03485	T-cell receptor be
74	64.6	201	2	T24183	hypothetical prote
75	64.6	261	2	C84366	probable oxidoredu
76	64.6	315	2	H86836	carbamate kinase (
77	64.6	327	2	T44432	fatty acid/phospho
78	64.6	339	2	E71918	uroporphyrinogen d
79	64.6	340	2	D64595	uroporphyrinogen d
80	64.6	351	1	A55962	opsin, pineal glan
81	64.6	370	2	I40358	N-acyl-L-amino aci
82	64.6	370	2	JC6130	paired box transcr
83	64.6	374	2	F69233	carbamoyl-phosphat
84	64.6	401	2	D83618	beta-ketoadipyl Co
85	64.6	422	2	AG1323	threonine dehydrat
86	64.6	422	2	AH1694	threonine dehydrat
87	64.6	434	2	AG1040	Vi polysaccharide
88	64.6	434	2	I36892	Vi polysaccharide
89	64.6	435	2	T35884	tagatose 6-phospha
90	64.6	455	2	B70664	hypothetical prote
91	64.6	521	2	S55317	cytochrome p450 1A
92	64.6	558	2	T19748	hypothetical prote
93	64.6	583	1	JC2545	acid phosphatase (
94	64.6	668	2	S56909	polymyxin B resist
95	64.6	772	2	S22860	excision repair pr
96	64.6	809	2	A55547	quinate-shikimate
97	64.6	964	2	S48404	probable membrane
98	64.6	1018	2	JC5799	alpha-D-mannosidas
99	64.6	1027	2	I38604	p53-binding protei
100	64.6	1039	2	T35878	hypothetical prote
101	64.6	1041	2	S42509	RAG-1 protein - ch
102	64.6	1042	2	S42511	RAG-1 protein - ra

103	31	64.6	1043	2	B33754	recombination-acti	176	30	62.5	813	2	S36018	gene tramtrack-p88
104	31	64.6	1043	2	A33754	recombination-acti	177	30	62.5	923	2	S44226	periodic tryptoph
105	31	64.6	1045	2	I51555	recombination acti	178	30	62.5	941	2	T33032	hypothetical prote
106	31	64.6	1087	2	F72283	beta-galactosidase	179	30	62.5	1157	2	T396761	hypothetical prote
107	31	64.6	1231	2	S53089	protein-tyrosine-p	180	30	62.5	1173	2	T30608	proteophosphoglyc
108	31	64.6	1254	2	T18277	kinesin heavy chai	181	30	62.5	1250	2	T00454	hypothetical prote
109	31	64.6	1488	2	T02856	probable membrane	182	30	62.5	1443	2	T31896	hypothetical prote
110	31	64.6	1762	2	T03222	probable polyketid	183	30	62.5	1587	2	A82012	hypothetical prote
111	31	64.6	15281	2	S41309	cyclosporin synth	184	30	62.5	1791	2	T02909	hypothetical prote
112	30.5	63.5	274	2	A84302	ketopantoate hydro	185	30	62.5	2318	2	S45306	notch 3 protein -
113	30	62.5	25	2	S22221	peroxidase (EC 1.1	186	30	62.5	3766	2	T29165	hypothetical prote
114	30	62.5	49	2	T07309	hypothetical prote	187	30	62.5	4848	2	T30289	pristinamycin I sy
115	30	62.5	114	2	E97017	hypothetical prote	188	30	62.5	26926	1	I38344	tiatin, cardiac mus
116	30	62.5	126	2	C90797	hypothetical prote	189	29	60.4	99	2	G83295	hypothetical prote
117	30	62.5	148	2	E87609	hypothetical prote	190	29	60.4	116	2	H50614	NADH dehydrogenase
118	30	62.5	196	2	T10206	hypothetical prote	191	29	60.4	122	2	A83427	transposase BMB114
119	30	62.5	210	2	D72490	hypothetical prote	192	29	60.4	130	2	T48771	hypothetical prote
120	30	62.5	212	2	B84337	deoxyribose-phosph	193	29	60.4	143	2	E69065	molybdopterin bios
121	30	62.5	225	2	F86417	hypothetical prote	194	29	60.4	145	2	D84776	probable ubiquitin
122	30	62.5	245	2	H97865	pseudouridylylate sy	195	29	60.4	146	2	T08443	probable DNA-bindi
123	30	62.5	262	1	FK8SK	kanamycin kinase (	196	29	60.4	154	2	A10347	probable exported
124	30	62.5	269	2	D82001	probable integral	197	29	60.4	159	2	T40564	hypothetical DNA-b
125	30	62.5	269	2	G81230	conserved hypotnet	198	29	60.4	174	1	WRBP70	early protein gp17
126	30	62.5	281	2	T35932	probable inositol	199	29	60.4	174	2	A86192	hypothetical prote
127	30	62.5	283	2	T15229	hypothetical prote	200	29	60.4	182	2	S78137	cytochrome-c oxida
128	30	62.5	286	2	T25559	hypothetical prote	201	29	60.4	184	2	B86192	hypothetical prote
129	30	62.5	291	2	S78602	peroxidase (EC 1.1	202	29	60.4	198	2	A83424	regulatory protein
130	30	62.5	296	2	T12469	hypothetical prote	203	29	60.4	200	2	C82224	probable alpha-rib
131	30	62.5	301	2	T37326	probable transcrip	204	29	60.4	205	2	C75096	probable acetyltra
132	30	62.5	312	2	E87705	hypothetical prote	205	29	60.4	210	1	JQ0394	nodB protein - Azo
133	30	62.5	319	2	B69806	divalent cation tr	206	29	60.4	221	2	B82647	hypothetical prote
134	30	62.5	331	2	C97493	malonyl-CoA-acyl c	207	29	60.4	234	2	H75390	hypothetical prote
135	30	62.5	331	2	A82711	malonyl-CoA-acyl c	208	29	60.4	240	2	G82295	conserved hypotnet
136	30	62.5	334	1	RDECEP	N-acetyl-gamma-glu	209	29	60.4	243	2	T36108	probable integral
137	30	62.5	334	2	G91239	N-acetyl-gamma-glu	210	29	60.4	245	2	B69118	conserved hypotnet
138	30	62.5	334	2	D86087	N-acetyl-gamma-glu	211	29	60.4	246	2	S27963	modulator recognit
139	30	62.5	351	2	D82930	SRP family of GTP-	212	29	60.4	249	2	F84118	hypothetical prote
140	30	62.5	360	2	C81220	hypothetical prote	213	29	60.4	263	2	B87486	hypothetical prote
141	30	62.5	363	2	S31780	peroxidase (EC 1.1	214	29	60.4	279	2	A80285	probable regulator
142	30	62.5	364	2	E81702	phospholipase D fa	215	29	60.4	289	2	F69313	conserved hypotnet
143	30	62.5	371	2	D85606	hypothetical prote	216	29	60.4	293	1	B9AG58	virB9 protein prec
144	30	62.5	389	2	C49776	hypothetical prote	217	29	60.4	293	2	A83249	component of type
145	30	62.5	393	2	AG0091	probable flagellar	218	29	60.4	301	2	E71558	probable fad synth
146	30	62.5	395	2	B47071	nitrogenase cofact	219	29	60.4	304	2	T40210	hypothetical prote
147	30	62.5	404	2	AE0973	O-antigen ligase I	220	29	60.4	313	2	A12806	carbohydrate kinas
148	30	62.5	404	2	B41317	O-antigen ligase c	221	29	60.4	313	2	H97585	carbohydrate kinas
149	30	62.5	405	2	T40193	hypothetical prote	222	29	60.4	315	2	H97368	truD protein (AE00
150	30	62.5	409	2	F90825	probable integrase	223	29	60.4	315	2	AH2586	trRNA pseudouridine
151	30	62.5	413	2	T48089	hypothetical prote	224	29	60.4	317	2	B81667	ferrochelatase TC0
152	30	62.5	416	2	A85684	probable integrase	225	29	60.4	322	2	H83852	L-asparaginase BH1
153	30	62.5	454	2	A82353	glutathione-disulf	226	29	60.4	324	2	T34196	hypothetical prote
154	30	62.5	483	2	S75369	hypothetical prote	227	29	60.4	324	2	E84918	hypothetical prote
155	30	62.5	493	2	C95216	galactose-1-phosph	228	29	60.4	325	2	T47229	arginase (EC 3.5.3
156	30	62.5	493	2	A98080	UTP-hexose-1-phosp	229	29	60.4	329	2	JC4093	signal recognition
157	30	62.5	497	2	D97264	galactose-1-phosph	230	29	60.4	330	2	F83960	signal recognition
158	30	62.5	497	2	S27785	acetyl-CoA acetyl	231	29	60.4	330	2	H95166	transcription regu
159	30	62.5	508	2	E83788	galactose-1-phosph	232	29	60.4	330	2	G98032	hypothetical prote
160	30	62.5	517	2	T05620	glycine hydroxymet	233	29	60.4	340	2	S75099	hypothetical prote
161	30	62.5	518	1	S40218	glycine hydroxymet	234	29	60.4	346	2	C30754	hypothetical prote
162	30	62.5	530	2	A45690	transactivator BEN	235	29	60.4	352	2	T36332	hypothetical prote
163	30	62.5	532	2	T01759	glycine hydroxymet	236	29	60.4	357	2	T20659	hypothetical prote
164	30	62.5	553	2	S77623	mannuronan C-5-epi	237	29	60.4	358	2	T47228	arginase (EC 3.5.3
165	30	62.5	561	2	G84012	oligo-1,6-glucosid	238	29	60.4	358	2	B64427	carbamoyl-phosphat
166	30	62.5	577	2	T18116	hypothetical prote	239	29	60.4	365	2	S56792	hypothetical prote
167	30	62.5	599	2	H72454	probable fatty-aci	240	29	60.4	365	2	F88449	protein F54D8.4 li
168	30	62.5	637	2	H71254	probable DNA gyrase	241	29	60.4	368	2	A11291	glycerol dehydroge
169	30	62.5	638	2	JC7753	ring finger B-box	242	29	60.4	368	2	AG1663	glycerol dehydroge
170	30	62.5	640	2	F90364	hypothetical prote	243	29	60.4	368	2	S75923	sensory transducti
171	30	62.5	642	2	S44542	probable membrane	244	29	60.4	382	2	A70705	probable PE protei
172	30	62.5	648	1	DUBPS2	DNA-directed DNA p	245	29	60.4	385	2	E82317	D-erythrose 4-phos
173	30	62.5	655	2	G70900	probable priA prot	246	29	60.4	391	2	G89822	hypothetical prote
174	30	62.5	718	2	A13420	penicillin-binding	247	29	60.4	393	2	H90291	thermostable carbo
175	30	62.5	739	1	VHIWEB	nucleocapsid prote	248	29	60.4	393	2	A99361	thermostable carbo

249	29	60.4	393	2	S76696	hypothetical prote	322	29	60.4	2204	1	RRNZV	genome polyprotein
250	29	60.4	393	2	AC1142	N-acyl-L-amino aci	323	29	60.4	2278	1	S56274	FAB1 protein - yea
251	29	60.4	393	2	AF1500	N-acyl-L-amino aci	324	29	60.4	2403	2	T30875	PRP8 protein homol
252	29	60.4	393	2	T49578	hypothetical prote	325	29	60.4	3172	2	S22012	erythronolide synt
253	29	60.4	394	2	S62726	translation elonga	326	29	60.4	3178	2	S13595	6-deoxyerythronoli
254	29	60.4	405	2	G86848	conserved hypotet	327	29	60.4	3588	2	I40485	surfactin syntheta
255	29	60.4	424	2	T40641	probable lipid met	328	29	60.4	4485	2	T08044	dynein gamma heav
256	29	60.4	426	2	H65102	probable tagatose	329	29	60.4	9376	2	T14593	synergomycin synth
257	29	60.4	426	2	H65102	probable tagatose	330	28.5	59.4	256	2	AH1978	hypothetical prote
258	29	60.4	426	2	B85975	probable tagatose	331	28.5	59.4	517	2	S40212	glycine hydroxymet
259	29	60.4	441	2	AH2068	N-acyl-L-amino aci	332	28.5	59.4	517	2	S40213	glycine hydroxymet
260	29	60.4	449	2	A43989	interferon-related	333	28.5	59.4	518	1	A42906	polyketide synthas
261	29	60.4	449	2	A43989	interferon-related	334	28.5	59.4	2257	2	T42918	hypothetical Bcl2/
262	29	60.4	455	1	T15581	phosphoprotein pho	335	28	58.3	27	4	I52725	probable exported
263	29	60.4	459	2	F71131	probable methyltra	336	28	58.3	70	2	AG0345	alpha-neurotoxin -
264	29	60.4	462	2	G70015	conserved hypotet	337	28	58.3	72	2	JC1474	gp68 protein - Myc
265	29	60.4	464	2	G70362	hypothetical prote	338	28	58.3	77	2	C72808	E7 protein - human
266	29	60.4	468	2	C82722	UMP-N-acetyluramo	339	28	58.3	86	2	S36533	carbon dioxide con
267	29	60.4	487	2	S61243	deoxyribonuclease	340	28	58.3	112	2	S77344	hypothetical prote
268	29	60.4	487	2	A87434	conserved hypotet	341	28	58.3	113	2	C69818	hypothetical prote
269	29	60.4	491	2	A24814	cytochrome P450 CY	342	28	58.3	113	2	D84389	hypothetical prote
270	29	60.4	491	2	A31418	cytochrome P450 2H	343	28	58.3	135	2	F72688	hypothetical prote
271	29	60.4	491	2	JC5312	UMP-hexose-1-phosp	344	28	58.3	148	2	AG2200	two-component syst
272	29	60.4	491	2	E95213	galactose-1-phosph	345	28	58.3	151	2	A39392	RAD6 DNA-repair ho
273	29	60.4	491	2	F98077	UMP-hexose-1-phosp	346	28	58.3	154	2	E84275	bacterioferritin c
274	29	60.4	508	2	H75605	hypothetical prote	347	28	58.3	179	2	S23358	H+-transporting tw
275	29	60.4	513	2	T45624	flavonoid 3'-hydro	348	28	58.3	181	2	AG1764	hypothetical prote
276	29	60.4	513	2	S39691	UMP-hexose-1-phosp	349	28	58.3	190	2	T21878	hypothetical prote
277	29	60.4	521	2	A29345	steroid hormone re	350	28	58.3	191	2	A82613	conserved hypotet
278	29	60.4	543	2	T06523	cytochrome P450 mo	351	28	58.3	192	2	T35904	probable transcrip
279	29	60.4	556	2	G86319	F25116.5 protein -	352	28	58.3	202	2	S59955	hypothetical prote
280	29	60.4	558	2	D69067	succinate dehydrog	353	28	58.3	204	2	G95017	hypothetical prote
281	29	60.4	568	2	S19031	poly(A) polymerase	354	28	58.3	204	2	H78990	conserved hypotet
282	29	60.4	577	2	B75621	sensor histidine k	355	28	58.3	211	2	E70045	two-component resp
283	29	60.4	590	2	T11098	NADH2 dehydrogenas	356	28	58.3	211	2	T35272	hypothetical prote
284	29	60.4	622	2	T27155	hypothetical prote	357	28	58.3	217	2	T37832	hypothetical prote
285	29	60.4	624	2	T01585	probable RING zinc	358	28	58.3	217	2	T03496	hypothetical prote
286	29	60.4	641	2	A42019	tyrosine-tRNA liga	359	28	58.3	219	2	S56810	conserved hypotet
287	29	60.4	651	2	S24609	cytoskeletal prote	360	28	58.3	221	2	H70395	conserved hypotet
288	29	60.4	665	1	T44793	beta-galactosidase	361	28	58.3	226	2	JQ2121	surface antigen -
289	29	60.4	684	2	T47694	probable serine/th	362	28	58.3	226	2	T27286	hypothetical prote
290	29	60.4	693	2	JN0673	ubiquitin-like fus	363	28	58.3	227	2	G83724	ABC transporter (A
291	29	60.4	700	2	D70951	probable UvrD - My	364	28	58.3	227	2	F86840	conserved hypotet
292	29	60.4	701	2	JN0674	ubiquitin-like fus	365	28	58.3	230	2	T11290	cytochrome-C oxida
293	29	60.4	733	2	A57459	ribosomal protein	366	28	58.3	230	2	D51686	hypothetical prote
294	29	60.4	770	2	S04847	leukocyte adhesio	367	28	58.3	232	2	D95113	O-methyltransferas
295	29	60.4	776	2	AH1506	preprotein translo	368	28	58.3	237	2	C97982	conserved hypotet
296	29	60.4	788	2	AB1991	hypothetical prote	369	28	58.3	237	2	D71287	hypothetical prote
297	29	60.4	867	2	T00118	hrSH2 protein - se	370	28	58.3	238	2	H84330	hypothetical prote
298	29	60.4	878	2	C71305	leucine-tRNA ligas	371	28	58.3	239	2	E84232	hypothetical prote
299	29	60.4	980	2	E86589	CR590 hypothetical	372	28	58.3	239	2	C84392	signal sequence pe
300	29	60.4	980	2	E72035	conserved hypotet	373	28	58.3	240	2	G89991	extracellular ente
301	29	60.4	1102	2	A84480	probable retroelem	374	28	58.3	255	2	D75415	conserved hypotet
302	29	60.4	1148	2	D85360	hypothetical prote	375	28	58.3	257	2	A71258	conserved hypotet
303	29	60.4	1209	2	E90994	probable regulator	376	28	58.3	262	2	E97151	ABC transporter MD
304	29	60.4	1209	2	H85839	probable regulator	377	28	58.3	265	2	JH0430	transformation com
305	29	60.4	1210	2	E64979	hypothetical 138.1	378	28	58.3	267	2	C83242	conserved hypotet
306	29	60.4	1243	2	T17390	vrK protein - Dic	379	28	58.3	267	2	E82318	transcription regu
307	29	60.4	1246	2	JQ0406	hypothetical prote	380	28	58.3	268	2	D81807	phosphomethylprim
308	29	60.4	1254	1	JQ1979	structural polypro	381	28	58.3	268	2	F81061	phosphomethylprim
309	29	60.4	1254	1	VHWVVE	structural polypro	382	28	58.3	269	2	S58439	transcription fact
310	29	60.4	1254	1	VHWVVT	structural polypro	383	28	58.3	272	2	F87351	hypothetical prote
311	29	60.4	1255	1	B44213	structural polypro	384	28	58.3	272	2	S61888	probable DNA-bindi
312	29	60.4	1255	1	D44213	structural polypro	385	28	58.3	277	2	B83207	probable transcrip
313	29	60.4	1282	2	JE0120	glycoprotein A - m	386	28	58.3	280	2	AF0649	formyltetrahydrofo
314	29	60.4	1283	2	T39174	hypothetical Serin	387	28	58.3	284	2	F71467	probable cell divi
315	29	60.4	1307	2	T30887	146D nuclear prote	388	28	58.3	289	2	A70604	hypothetical prote
316	29	60.4	1322	2	T24140	hypothetical prote	389	28	58.3	294	2	A72362	cell division prot
317	29	60.4	1354	2	T13930	tripeptidyl-peptid	390	28	58.3	297	2	T34165	hypothetical prote
318	29	60.4	1364	2	T00250	MEGF2 protein - hu	391	28	58.3	303	2	S76535	probable phosphoe
319	29	60.4	1466	2	G84516	probable retroelem	392	28	58.3	306	2	D95273	conserved hypotet
320	29	60.4	1506	2	S52957	bIMD protein - Eme	393	28	58.3	308	2	G69135	hypothetical prote
321	29	60.4	1974	2	T16703	hypothetical prote	394	28	58.3	309	2	T36244	hypothetical prote

395	28	58.3	323	2	I49529	transcription fact	468	28	58.3	484	2	AF2221	hypothetical prote
396	28	58.3	327	2	E84137	rhizopine ABC tran	469	28	58.3	488	2	S37466	sucrose phosphoryl
397	28	58.3	329	2	T10444	peroxidase [EC 1.1	470	28	58.3	492	2	A53040	conserved hypotet
398	28	58.3	329	2	T22221	hypothetical prote	471	28	58.3	505	2	D70782	hypothetical prote
399	28	58.3	332	2	F83600	conserved hypotet	472	28	58.3	509	2	B98246	hypothetical prote
400	28	58.3	334	2	F75344	probable polyferre	473	28	58.3	512	2	T47554	cytochrome P450 ho
401	28	58.3	335	2	JC1441	transcription fact	474	28	58.3	515	2	T36714	probable protein p
402	28	58.3	340	2	H83805	NAAD-dependent dyh	475	28	58.3	517	2	B70687	hypothetical prote
403	28	58.3	341	2	S09913	hypothetical prote	476	28	58.3	533	2	S76101	hypothetical prote
404	28	58.3	345	2	B75274	conserved hypotet	477	28	58.3	538	2	F69215	adenine deaminase
405	28	58.3	349	2	T24308	hypothetical prote	478	28	58.3	544	2	I55454	neuroglycan C prec
406	28	58.3	351	2	AF2844	hypothetical prote	479	28	58.3	547	2	AE1022	probable membrane
407	28	58.3	351	2	H97621	lipopolysaccharide	480	28	58.3	547	2	A40656	hypothetical prote
408	28	58.3	353	2	H84332	carbamoyl-phosphat	481	28	58.3	553	1	G1BPSV	gene 1 protein - s
409	28	58.3	353	2	A95998	probable sugar upt	482	28	58.3	553	2	S76949	hypothetical prote
410	28	58.3	358	2	H69408	carbamoyl-phosphat	483	28	58.3	559	2	AC3373	NAD synthase (Gluc
411	28	58.3	359	2	B96505	probable gibbereli	484	28	58.3	566	2	T31964	hypothetical prote
412	28	58.3	362	2	A89777	capsular polysacch	485	28	58.3	570	2	G85046	hypothetical prote
413	28	58.3	373	2	H72756	probable phosphate	486	28	58.3	582	2	A70755	hypothetical prote
414	28	58.3	374	2	T21513	hypothetical prote	487	28	58.3	586	2	B70503	probable pyrG prot
415	28	58.3	375	2	AH0306	probable exported	488	28	58.3	593	2	F95939	probable adenine d
416	28	58.3	375	2	H71121	probable alcohol d	489	28	58.3	594	2	AB3343	single-stranded-DN
417	28	58.3	377	2	F71877	alanine racemase -	490	28	58.3	595	2	T26843	hypothetical prote
418	28	58.3	382	2	T28691	hypothetical prote	491	28	58.3	597	2	D98186	adenine deaminase
419	28	58.3	384	2	A72257	probable transamin	492	28	58.3	604	2	AF3100	probable centromer
420	28	58.3	388	2	D75496	aspartate transami	493	28	58.3	614	2	T39740	probable beta-gluc
421	28	58.3	388	2	S41940	oligogalacturonide	494	28	58.3	619	2	T01121	72K mitochondrial
422	28	58.3	388	2	JQ0189	oligogalacturonide	495	28	58.3	627	2	A36682	hypothetical prote
423	28	58.3	388	2	AI0208	oligogalacturonide	496	28	58.3	627	2	S46820	hypothetical prote
424	28	58.3	391	1	TVBE11	44K protein kinase	497	28	58.3	631	2	T48255	hypothetical prote
425	28	58.3	392	2	T48133	hypothetical prote	498	28	58.3	663	2	S55164	hypothetical prote
426	28	58.3	395	2	AB2977	hippurate hydrolas	499	28	58.3	686	1	ALDYAT	amylase A (EC 3.2.
427	28	58.3	396	2	A75345	probable chloromuc	500	28	58.3	687	2	S51398	hypothetical prote
428	28	58.3	399	2	D95279	probable alcohol f	501	28	58.3	711	2	A90023	DNA topoisomerase
429	28	58.3	399	2	AE3556	alanine racemase (	502	28	58.3	712	1	BVRCB	CyAB protein - Bor
430	28	58.3	400	2	T35152	hypothetical prote	503	28	58.3	731	1	A55800	cucumisin (EC 3.4.
431	28	58.3	402	2	H64599	poly(A) polymerase	504	28	58.3	739	1	JQ1893	80.7K alpha trans-
432	28	58.3	402	2	E71912	polynucleotide ade	505	28	58.3	741	2	D95966	probable aldehyde
433	28	58.3	404	2	JC5784	adrenomedullin rec	506	28	58.3	741	2	T40095	Zinc finger, C3HC4
434	28	58.3	407	2	S25841	beta-ketoacyl synt	507	28	58.3	742	1	TNBEH1	80.7K alpha trans-
435	28	58.3	408	2	C95100	aminotransferase,	508	28	58.3	750	2	C87159	cation-transportin
436	28	58.3	408	2	E97968	conserved hypotet	509	28	58.3	750	2	S77653	cation-transportin
437	28	58.3	408	2	AD3237	conserved hypotet	510	28	58.3	753	2	T52360	hypothetical prote
438	28	58.3	416	2	F24723	tryptophan synthas	511	28	58.3	784	2	E82731	outer membrane ant
439	28	58.3	416	2	S76858	hypothetical prote	512	28	58.3	789	2	T49055	hypothetical prote
440	28	58.3	419	2	A85864	hypothetical prote	513	28	58.3	796	2	JC7516	vesicle protein so
441	28	58.3	419	2	G91019	hypothetical prote	514	28	58.3	820	2	T32908	hypothetical prote
442	28	58.3	419	2	S56073	opaque-2 protein -	515	28	58.3	845	2	T25657	alanine-tRNA ligas
443	28	58.3	422	2	H82236	iron-containing al	516	28	58.3	875	2	A80401	heme transport pro
444	28	58.3	422	2	A98306	probable hydrolase	517	28	58.3	877	2	AC2211	antibiotic synthet
445	28	58.3	423	2	T36467	probable glycosyl	518	28	58.3	879	2	B70014	protein-histidine
446	28	58.3	425	1	JH0710	transcription fact	519	28	58.3	904	2	G64840	sensor protein cor
447	28	58.3	428	2	I51087	G protein-coupled	520	28	58.3	904	2	D90772	sensor protein cor
448	28	58.3	430	2	D75382	isocitrate dehydro	521	28	58.3	904	2	H85634	N-methyl-D-asparta
449	28	58.3	433	2	C83271	conserved hypotet	522	28	58.3	965	2	I51244	probable ABC-type
450	28	58.3	434	2	E81139	hypothetical prote	523	28	58.3	1011	2	T07712	major tegumental a
451	28	58.3	440	2	C83368	probable NFS trans	524	28	58.3	1032	2	T30270	hypothetical prote
452	28	58.3	445	2	AD0962	DsdX permease [imp	525	28	58.3	1034	2	S76134	integrin alpha-4 c
453	28	58.3	449	2	F70526	probable cytochrom	526	28	58.3	1038	2	S06046	lysine-ketoglutar
454	28	58.3	450	2	B69198	UDP-N-acetylmuram	527	28	58.3	1056	2	T02930	protein-tyrosine k
455	28	58.3	456	2	AF0093	Sodium sulfate sym	528	28	58.3	1122	2	I54237	protein-tyrosine k
456	28	58.3	458	2	F72244	conserved hypotet	529	28	58.3	1123	1	JN0712	protein-tyrosine k
457	28	58.3	461	2	I64080	probable membrane	530	28	58.3	1124	1	I58388	protein-tyrosine k
458	28	58.3	465	1	DEBY4	alcohol dehydrogen	531	28	58.3	1125	1	JH0771	protein-tyrosine k
459	28	58.3	465	2	F70364	dihydroipoamide d	532	28	58.3	1125	1	S77846	hypothetical prote
460	28	58.3	467	2	T25848	hypothetical prote	533	28	58.3	1130	2	T23104	hypothetical prote
461	28	58.3	467	2	D85683	ubiquinone oxidore	534	28	58.3	1133	2	T23103	hypothetical prote
462	28	58.3	467	2	F72040	probable sodium-tr	535	28	58.3	1180	2	T31066	vascular cadherin-
463	28	58.3	467	2	B82225	protein T12M4.6 li	536	28	58.3	1194	1	G70837	probable ABC trans
464	28	58.3	468	2	S39832	probable phosphopr	537	28	58.3	1201	2	F81202	proline dehydrogen
465	28	58.3	469	2	F76891	hypothetical prote	538	28	58.3	1203	2	C95229	DNA-directed RNA p
466	28	58.3	475	2	F85430	serine C-palmitoyl	539	28	58.3	1216	2	G98093	DNA-directed RNA p
467	28	58.3	480	2	G70678	hypothetical prote	540	28	58.3	1218	2	AC3580	1-pyrroline-5-carb

541	28	58.3	1224	2	T43218	proline dehydrogen	614	27	56.2	160	2	S45632	H+-transporting tw
542	28	58.3	1228	2	C98219	proline dehydrogen	615	27	56.2	161	2	T09807	probable glutathio
543	28	58.3	1228	2	AG3067	proline dehydrogen	616	27	56.2	166	2	G72641	hypothetical prote
544	28	58.3	1230	2	H84515	probable helicase	617	27	56.2	166	2	G83384	probable ring-hydr
545	28	58.3	1274	2	E81779	proline dehydrogen	618	27	56.2	172	2	T27505	hypothetical prote
546	28	58.3	1275	2	T18556	O-antigen biosynth	619	27	56.2	174	2	F83581	heme d1 biosynthes
547	28	58.3	1303	2	C87519	hypothetical prote	620	27	56.2	183	2	S53149	core antigen - hep
548	28	58.3	1310	1	I53597	proline dehydrogen	621	27	56.2	183	2	T38917	hypothetical prote
549	28	58.3	1310	1	D64843	proline dehydrogen	622	27	56.2	190	2	E84060	hypothetical prote
550	28	58.3	1320	1	S66279	proline dehydrogen	623	27	56.2	191	2	B84272	protoporphyrinogen
551	28	58.3	1320	2	AE0633	proline dehydrogen	624	27	56.2	191	2	S49826	hypothetical prote
552	28	58.3	1320	2	D90786	proline dehydrogen	625	27	56.2	201	2	S86488	hypothetical prote
553	28	58.3	1320	2	B85646	proline dehydrogen	626	27	56.2	203	2	D83236	probable sigma-70
554	28	58.3	1323	2	AH0225	1-pyrroline-5-carb	627	27	56.2	206	2	B61121	hypothetical prote
555	28	58.3	1330	2	T30341	zinc finger protei	628	27	56.2	208	2	T10537	hypothetical prote
556	28	58.3	1386	2	T18434	hypothetical prote	629	27	56.2	210	2	F83751	ABC transporter (A
557	28	58.3	1403	1	A47328	natural killer cel	630	27	56.2	216	1	BPSPBP	gldB protein - Pse
558	28	58.3	1535	2	S46224	peroxidase - frui	631	27	56.2	216	2	I38482	olfactory receptor
559	28	58.3	1617	2	B86483	protein F5U5.15 [i	632	27	56.2	216	2	AH2547	hypothetical prote
560	28	58.3	1636	2	B82736	hemolysin-type cal	633	27	56.2	218	2	S76385	hypothetical prote
561	28	58.3	1640	2	H88094	protein F39E9.2 [i	634	27	56.2	220	2	T34986	hypothetical prote
562	28	58.3	1715	2	G84429	hypothetical prote	635	27	56.2	224	1	RGEBFT	probable secreted
563	28	58.3	1848	2	A44140	cellulose-binding	636	27	56.2	224	2	AH0646	transcription regu
564	28	58.3	1896	2	T08851	Down syndrome cell	637	27	56.2	228	2	T46859	lexA repressor pro
565	28	58.3	1957	2	T38077	hypothetical coile	638	27	56.2	229	2	T11104	transcription regu
566	28	58.3	2052	2	T18519	myosin X - bovine	639	27	56.2	229	2	G86924	cytochrome-c oxida
567	28	58.3	2185	1	S60200	acetyl-CoA carboxy	640	27	56.2	229	2	D87221	hypothetical prote
568	28	58.3	2281	2	T07084	acetyl-CoA carboxy	641	27	56.2	239	2	B82123	Hsp70 cofactor [im
569	28	58.3	2453	2	S60254	nuclear receptor c	642	27	56.2	239	2	S64573	chemotaxis protein
570	28	58.3	2607	2	T31678	bacitracin synthet	643	27	56.2	239	2	A99291	hypothetical prote
571	28	58.3	3573	2	S23070	erythronolide synt	644	27	56.2	240	2	AC2748	hypothetical prote
572	28	58.3	3643	2	T36410	probable polyketid	645	27	56.2	240	2	B97529	lexA repressor [im
573	28	58.3	4845	2	T31067	BIR repeat contain	646	27	56.2	240	2	AB3357	hypothetical prote
574	28	58.3	5175	2	T20992	hypothetical prote	647	27	56.2	242	2	E83738	hypothetical prote
575	28	58.3	5198	2	T43290	hemikentlin precurs	648	27	56.2	244	2	T48110	repressor LexA [im
576	28	58.3	5198	2	B89921	hypothetical prote	649	27	56.2	245	2	AH7485	hypothetical prote
577	27.5	57.3	430	2	T28143	tapasin 1 homolog,	650	27	56.2	245	2	D70670	photosystem II oxy
578	27.5	57.3	469	2	S61632	glycine hydroxymet	651	27	56.2	248	2	S03888	ABC transporter, A
579	27	56.2	40	2	S33290	lipopolysaccharide	652	27	56.2	248	2	E75302	transcription regu
580	27	56.2	51	2	T37012	hypothetical prote	653	27	56.2	249	2	AF1321	phosphate transpor
581	27	56.2	62	2	F81810	hypothetical prote	654	27	56.2	253	2	E69098	hypothetical prote
582	27	56.2	65	2	F83727	hypothetical prote	655	27	56.2	253	2	T29704	photosystem II oxy
583	27	56.2	66	2	A87561	hypothetical prote	656	27	56.2	254	2	T02078	probable photosyst
584	27	56.2	79	2	S31013	gene 68 protein -	657	27	56.2	254	2	T02873	photosystem II oxy
585	27	56.2	94	2	H83655	hypothetical prote	658	27	56.2	256	2	B82076	probable general s
586	27	56.2	99	2	B97409	hypothetical prote	659	27	56.2	257	2	G84712	hypothetical prote
587	27	56.2	104	2	T23203	hypothetical prote	660	27	56.2	258	1	F2TOX2	photosystem II oxy
588	27	56.2	107	2	T23204	hypothetical prote	661	27	56.2	258	2	S22763	photosystem II oxy
589	27	56.2	108	1	CCWB	cytochrome c - ear	662	27	56.2	258	2	B87615	ThiJ/FfpI family p
590	27	56.2	112	1	ASLJTX	vpv protein - huma	663	27	56.2	259	2	JS0771	photosystem II oxy
591	27	56.2	113	2	F64636	wbeB protein homol	664	27	56.2	259	2	D84343	TRK potassium upta
592	27	56.2	116	2	A10746	flagellar transcri	665	27	56.2	260	2	S10016	photosystem II oxy
593	27	56.2	119	1	XMECFB	flagellar transcri	666	27	56.2	260	2	G82639	hypothetical prote
594	27	56.2	119	2	B90954	regulator of flagel	667	27	56.2	261	2	A90926	hypothetical prote
595	27	56.2	119	2	F85802	regulator of flagel	668	27	56.2	261	2	B85774	hypothetical prote
596	27	56.2	121	2	S59680	ribosomal protein	669	27	56.2	261	2	F64924	probable thiosulfa
597	27	56.2	125	2	G71000	hypothetical prote	670	27	56.2	263	2	JC5271	oxygen-evolving co
598	27	56.2	128	2	G97176	stress-induced pro	671	27	56.2	263	2	T02227	NBS-LRR type resis
599	27	56.2	129	2	FC1254	endonuclease (BC 3	672	27	56.2	263	2	A83705	phosphonates trans
600	27	56.2	131	2	AD2648	conserved hypoteth	673	27	56.2	264	2	T35168	probable transcrip
601	27	56.2	131	2	T36638	probable substrate	674	27	56.2	266	2	A89775	hypothetical prote
602	27	56.2	133	2	C97430	hypothetical prote	675	27	56.2	266	2	T44344	late competence pr
603	27	56.2	137	1	S77065	hypothetical prote	676	27	56.2	271	2	C47127	tetracenomycin C s
604	27	56.2	138	2	AD2050	hypothetical prote	677	27	56.2	272	2	H90293	hypothetical prote
605	27	56.2	139	2	D82286	ferredoxin VC0753	678	27	56.2	273	2	D84195	oxidoeductase [im
606	27	56.2	141	2	T49659	hypothetical prote	679	27	56.2	273	2	G87037	conserved hypoteth
607	27	56.2	142	2	B70634	probable mmpS1 pro	680	27	56.2	274	2	AF3291	hypothetical prote
608	27	56.2	144	2	E72705	hypothetical prote	681	27	56.2	274	2	AE3552	high-affinity bran
609	27	56.2	145	2	T34303	hypothetical prote	682	27	56.2	282	2	C71444	probable thioester
610	27	56.2	147	2	A42435	leech antiplatelet	683	27	56.2	283	2	S10773	2-hydroxyuconic s
611	27	56.2	157	2	D70777	hypothetical prote	684	27	56.2	283	2	F95893	probable ABC trans
612	27	56.2	157	2	F87574	CBS domain protein	685	27	56.2	284	2	F81729	signal recognition
613	27	56.2	158	2	T27757	hypothetical prote	686	27	56.2	286	2	B87541	diene lactone hydro

687	27	56.2	286	2	B70833	carbon-monoxide de	760	27	56.2	373	2	B26470	spore germination
688	27	56.2	286	2	B96615	probable carbonic	761	27	56.2	374	2	AH0908	probable ATP/GTP-b
689	27	56.2	287	2	A81803	probable integral	762	27	56.2	375	2	C75040	alcohol dehydrogen
690	27	56.2	288	2	D81065	hypothetical prote	763	27	56.2	375	2	A91142	hypothetical prote
691	27	56.2	291	2	C97453	cytochrome c oxida	764	27	56.2	375	2	B65115	hypothetical 43.1
692	27	56.2	291	2	AE2671	cytochrome c oxida	765	27	56.2	375	2	D85987	hypothetical prote
693	27	56.2	291	2	H75512	formamidopyrimidin	766	27	56.2	380	2	B86625	DNA-directed DNA p
694	27	56.2	291	2	A38051	H+/K+-exchanging A	767	27	56.2	380	2	T30306	DNA-directed DNA p
695	27	56.2	294	2	G72293	phosphate butyrylt	768	27	56.2	385	2	T26638	hypothetical prote
696	27	56.2	295	2	A84145	hypothetical prote	769	27	56.2	388	2	F97331	alpha/beta superfa
697	27	56.2	300	1	NDBPT7	exodeoxyribonuclea	770	27	56.2	391	2	H86315	hypothetical prote
698	27	56.2	300	2	G86829	ribokinase (EC 2.7	771	27	56.2	391	2	C94166	hypothetical prote
699	27	56.2	300	2	AB3106	hypothetical prote	772	27	56.2	391	2	T08264	sodium-dependent p
700	27	56.2	301	2	P86749	pseudouridine synt	773	27	56.2	398	2	S16275	naringenin-chalcon
701	27	56.2	302	2	T36030	probable fructokin	774	27	56.2	398	2	C64764	membrane protein y
702	27	56.2	302	2	AG3407	florfenicol resist	775	27	56.2	400	2	S70187	44.7K virA protein
703	27	56.2	304	2	AB2663	2-dehydro-3-deoxyg	776	27	56.2	402	2	C96781	unknown protein F9
704	27	56.2	305	2	T44955	ribokinase (EC 2.7	777	27	56.2	403	2	C33958	hypothetical prote
705	27	56.2	306	2	A25698	probable protein k	778	27	56.2	405	2	AF2422	N-acyl-L-amino aci
706	27	56.2	307	2	A64449	formylmethanofuran	779	27	56.2	405	2	B36807	hypothetical prote
707	27	56.2	307	2	G75631	probable iron-chel	780	27	56.2	405	2	AH2752	dihydrolipoamide a
708	27	56.2	308	1	QQLJX1	trans-activating t	781	27	56.2	408	2	PD0007	cytochrome P450 no
709	27	56.2	309	2	B87712	conserved hypothet	782	27	56.2	408	2	JC5674	RAD23 protein homo
710	27	56.2	311	2	G86383	probable mitochond	783	27	56.2	409	2	S44346	conserved hypothet
711	27	56.2	314	2	A56650	2-oxoglutarate car	784	27	56.2	418	2	H82209	CTD11 hypothetical
712	27	56.2	314	2	A36305	2-oxoglutarate/mal	785	27	56.2	420	2	B86503	CTD11 hypothetical
713	27	56.2	314	2	H70723	hypothetical prote	786	27	56.2	420	2	D72119	CTD11 hypothetical
714	27	56.2	315	2	S11081	Na+/K+-exchanging	787	27	56.2	421	2	H81243	signal recognition
715	27	56.2	315	2	B98181	sitB protein (AFL12	788	27	56.2	421	2	T44429	probable transcrip
716	27	56.2	316	2	H95985	probable transcrip	789	27	56.2	421	2	F82024	probable signal re
717	27	56.2	322	2	H83851	hypothetical prote	790	27	56.2	422	2	G90259	hypothetical prote
718	27	56.2	323	2	T25459	hypothetical prote	791	27	56.2	422	2	H90275	hypothetical prote
719	27	56.2	323	2	A44504	suHk protein - Rhl	792	27	56.2	422	2	B90327	hypothetical prote
720	27	56.2	323	2	S19994	acid phosphatase (	793	27	56.2	424	2	A30474	hypothetical prote
721	27	56.2	324	2	B25102	reaction center pr	794	27	56.2	425	2	C81551	conserved hypothet
722	27	56.2	325	2	H81553	hypothetical prote	795	27	56.2	425	2	T72741	hypothetical prote
723	27	56.2	325	2	D86507	hypothetical prote	796	27	56.2	427	2	T05019	hypothetical prote
724	27	56.2	325	2	E72116	hypothetical prote	797	27	56.2	430	2	F89923	asparaginyl-tRNA s
725	27	56.2	325	2	T10045	hypothetical prote	798	27	56.2	430	2	B89591	asparaginyl-tRNA li
726	27	56.2	325	2	E70684	hypothetical prote	799	27	56.2	430	2	H83861	asparaginyl-tRNA s
727	27	56.2	325	2	A40963	phospholipase A2-a	800	27	56.2	430	2	AH1311	asparaginyl-tRNA s
728	27	56.2	326	2	S61517	ficollin-1 precuro	801	27	56.2	430	2	AH1683	asparaginyl-tRNA s
729	27	56.2	329	2	A97445	hypothetical prote	802	27	56.2	430	2	C97160	uracil permease ur
730	27	56.2	333	2	B86179	hypothetical prote	803	27	56.2	430	2	T21060	hypothetical prote
731	27	56.2	340	2	B89031	conserved hypothet	804	27	56.2	431	2	E81053	seryl-tRNA synthet
732	27	56.2	341	2	AH3016	fructose biphosph	805	27	56.2	435	2	B82135	menaquinone-specif
733	27	56.2	341	2	H98267	hypothetical prote	806	27	56.2	436	2	B83747	dihydrolipoamide S
734	27	56.2	341	2	A64383	hypothetical prote	807	27	56.2	436	2	H69588	acetylornithine de
735	27	56.2	345	2	E84398	phosphate ABC tran	808	27	56.2	437	2	A38534	branched-chain ami
736	27	56.2	345	2	A43586	hypothetical prote	809	27	56.2	438	2	I38946	melanoma ubiquitou
737	27	56.2	345	2	G63138	conserved hypothet	810	27	56.2	439	2	H69783	conserved hypothet
738	27	56.2	345	2	AH1514	hypothetical prote	811	27	56.2	447	2	AB3359	dihydrolipoamide S
739	27	56.2	347	2	B87500	conserved hypothet	812	27	56.2	448	2	AF2198	AAA superfamily AT
740	27	56.2	348	2	T22065	hypothetical prote	813	27	56.2	451	2	T16000	hypothetical prote
741	27	56.2	349	2	B81736	probable 3'(2'),5'	814	27	56.2	456	2	G81660	Glpr/Pept/Uhpf fam
742	27	56.2	351	2	T32717	hypothetical prote	815	27	56.2	456	2	A71501	probable hexesphos
743	27	56.2	352	1	BVECHD	molybdenum transpo	816	27	56.2	457	2	F37533	dihydrolipoamide a
744	27	56.2	352	2	AE0595	molybdenum transpo	817	27	56.2	465	2	AG0415	aromatic amino aci
745	27	56.2	352	2	B85579	ATP-binding compon	818	27	56.2	470	2	G87085	arginosuccinate ly
746	27	56.2	352	2	A90728	ATP-binding compon	819	27	56.2	470	2	S78440	phosphoglucomutase
747	27	56.2	352	2	C69401	conserved hypothet	820	27	56.2	473	2	A55591	isocitrate dehydro
748	27	56.2	354	2	H82557	hypothetical prote	821	27	56.2	473	2	AE2034	isocitrate dehydro
749	27	56.2	358	2	H89954	hypothetical prote	822	27	56.2	475	2	S74684	probable prolycar
750	27	56.2	358	2	AH7514	hypothetical prote	823	27	56.2	476	2	G84634	probable transport
751	27	56.2	359	2	S27788	neutrophil oxidase	824	27	56.2	476	2	A83387	hypothetical prote
752	27	56.2	360	2	E84413	geranylgeranyl hyd	825	27	56.2	478	2	T32825	hypothetical prote
753	27	56.2	362	2	A86831	X-Pro dipeptidase	826	27	56.2	488	2	AD0679	fructuronate reduc
754	27	56.2	364	2	T10945	peroxidase (EC 1.1	827	27	56.2	490	2	I48163	cytochrome P450 -
755	27	56.2	366	2	JC4104	protidine-5'-phosp	828	27	56.2	490	2	AE2050	signal recognition
756	27	56.2	367	2	JC4103	protidine-5'-phosp	829	27	56.2	492	2	F75389	NADH2 dehydrogenas
757	27	56.2	367	2	AC2469	hypothetical prote	830	27	56.2	494	2	F83206	probable aldehyde
758	27	56.2	372	2	C81296	carbamoyl-phosphat	831	27	56.2	496	2	G89824	hypothetical prote
759	27	56.2	373	2	G69629	germination respon	832	27	56.2	507	2	A39411	glucose-6-phosphat

833	27	56.2	509	2	E89787	hypothetical prote	906	27	56.2	802	2	F70600	hypothetical prote
834	27	56.2	509	2	JC5651	N-acetylglucosamin	907	27	56.2	813	2	JC5659	translation initia
835	27	56.2	510	2	A86143	probable cytochrom	908	27	56.2	822	2	S68210	chloride channel p
836	27	56.2	510	2	C29782	long-chain fatty a	909	27	56.2	832	2	S76815	hypothetical prote
837	27	56.2	510	2	C97561	hypothetical prote	910	27	56.2	836	2	S54152	sepB protein - Eme
838	27	56.2	515	2	G75267	ABC transporter, p	911	27	56.2	842	2	T04880	potassium transpor
839	27	56.2	516	1	D34306	phenol 2-monooxyge	912	27	56.2	859	2	T35785	probable beta-gluc
840	27	56.2	517	1	D37831	PTS system, N-acet	913	27	56.2	859	2	T05470	hypothetical prote
841	27	56.2	523	2	F82254	adenylosuccinate l	914	27	56.2	876	1	SHNC	phosphoribosyl-AMP
842	27	56.2	532	2	D82666	I2 protein - human	915	27	56.2	876	2	T33176	hypothetical prote
843	27	56.2	533	2	S36594	pyruvate dehydrog	916	27	56.2	879	2	T49796	probable sepB prot
844	27	56.2	538	2	A48692	probable membrane	917	27	56.2	880	2	AC2108	alanyl-CRNA synthe
845	27	56.2	545	2	AB0290	60K cysteine-rich	918	27	56.2	885	1	S26723	DNA-directed RNA p
846	27	56.2	547	1	A32244	60K cysteine-rich	919	27	56.2	887	2	A54832	villin homolog qua
847	27	56.2	547	2	B43584	alpha, alpha-phosph	920	27	56.2	891	2	AC0149	DNA topoisomerase
848	27	56.2	548	2	AF1231	alpha, alpha-phosph	921	27	56.2	901	2	H64101	preprotein translo
849	27	56.2	548	2	AE1585	chaperonin - Metha	922	27	56.2	907	2	S23399	chloride channel p
850	27	56.2	552	2	H69126	60K cysteine-rich	923	27	56.2	917	2	S51254	STE5 protein - yea
851	27	56.2	553	2	D71515	hypothetical prote	924	27	56.2	918	2	S04255	regulatory protein
852	27	56.2	556	2	S75873	urocanate hydratase	925	27	56.2	919	2	C87445	DNA gyrase subunit
853	27	56.2	558	2	B75611	phosphorylase kina	926	27	56.2	920	2	C70668	probable mmpL7 pro
854	27	56.2	560	2	S51600	hypothetical prote	927	27	56.2	966	2	S18955	fix23-3 protein -
855	27	56.2	562	2	H70888	integral membrane	928	27	56.2	987	2	A75496	conserved hypothet
856	27	56.2	564	2	S75777	alkaline phosphata	929	27	56.2	987	2	I48373	G-utrophin - mouse
857	27	56.2	564	2	F87305	probable membrane	930	27	56.2	1002	2	T30546	major surface glyc
858	27	56.2	567	2	AC0754	DNA repair protein	931	27	56.2	1046	2	S67786	hypothetical prote
859	27	56.2	568	2	B87495	acetolactate synth	932	27	56.2	1059	1	A35210	ferroxidase (EC 1.
860	27	56.2	571	2	F84956	hypothetical prote	933	27	56.2	1062	2	S61196	SUM1 protein - yea
861	27	56.2	571	2	G84426	pyruvate dehydrog	934	27	56.2	1063	2	T00624	endo-1,4-beta-xyla
862	27	56.2	572	2	AG0608	probable oxidoredu	935	27	56.2	1076	2	B85295	probable DNA misem
863	27	56.2	574	2	C83359	hypothetical prote	936	27	56.2	1076	2	T05793	mah3 protein homol
864	27	56.2	577	2	E71364	hypothetical prote	937	27	56.2	1078	2	PC4198	peptide synthetase
865	27	56.2	578	2	T19023	hypothetical prote	938	27	56.2	1081	2	T51613	DNA mismatch repai
866	27	56.2	582	2	F82064	conserved hypothet	939	27	56.2	1121	2	T47970	hypothetical prote
867	27	56.2	588	2	B87521	peptidase M1 famil	940	27	56.2	1148	2	F69685	pyruvate carboxyla
868	27	56.2	590	1	A54372	G protein-coupled	941	27	56.2	1159	2	T13946	probable adaptor-r
869	27	56.2	596	2	G69038	dnak-type molecule	942	27	56.2	1225	2	B95229	DNA-directed RNA p
870	27	56.2	598	2	T48822	hypothetical prote	943	27	56.2	1225	2	F98093	DNA-binding protei
871	27	56.2	599	2	S18735	centromere protein	944	27	56.2	1235	1	Q0BEW4	structural polypro
872	27	56.2	602	2	T21980	hypothetical prote	945	27	56.2	1254	1	JQ1978	hypothetical prote
873	27	56.2	607	2	S96598	DNA polymerase III	946	27	56.2	1254	2	T47141	hypothetical prote
874	27	56.2	608	2	B87282	hypothetical prote	947	27	56.2	1289	2	S69689	hypothetical prote
875	27	56.2	610	2	G72494	ABC transporter, A	948	27	56.2	1291	2	T17242	multidrug resistan
876	27	56.2	619	2	G75321	carbon-monoxide de	949	27	56.2	1302	2	B41249	cell division prot
877	27	56.2	622	1	H69480	phosphorotripho	950	27	56.2	1329	2	A64828	epidermal growth f
878	27	56.2	623	2	S68963	ABC transporter AT	951	27	56.2	1330	1	GQFFE	cell division prot
879	27	56.2	626	2	A13310	hypothetical prote	952	27	56.2	1342	2	B85614	cell division prot
880	27	56.2	628	2	T20714	conserved hypothet	953	27	56.2	1342	2	G90750	cell division prot
881	27	56.2	628	2	H89917	probable polygalac	954	27	56.2	1343	2	AF0611	hypothetical prote
882	27	56.2	629	2	T07426	hypothetical prote	955	27	56.2	1374	2	D72593	hypothetical prote
883	27	56.2	629	2	T25571	ribosomal protein	956	27	56.2	1385	2	T25828	hypothetical prote
884	27	56.2	633	2	C32571	outer membrane pro	957	27	56.2	1441	2	B86807	nuclear pore compl
885	27	56.2	634	2	A64521	hypothetical prote	958	27	56.2	1475	2	S42718	hypothetical prote
886	27	56.2	638	2	T48380	O-succinylhomoseri	959	27	56.2	1533	2	F71274	ClaHtrn heavy cha
887	27	56.2	639	2	S57153	probable penicilli	960	27	56.2	1681	2	S42369	serine/threonine k
888	27	56.2	652	2	C71285	probable outer mem	961	27	56.2	1799	2	AD1895	acetyl-CoA carboxy
889	27	56.2	668	2	A71986	harpin secretion	962	27	56.2	1822	2	T02235	hypothetical prote
890	27	56.2	697	2	A49313	conserved hypothet	963	27	56.2	2325	2	T50207	hypothetical prote
891	27	56.2	705	2	S55420	chemotaxis protein	964	27	56.2	2374	2	T21052	inositol 1,4,5-tri
892	27	56.2	720	2	C82380	hypothetical prote	965	27	56.2	2783	2	T31431	364K Golgi complex
893	27	56.2	720	2	S75935	ribosomal protein	966	27	56.2	3187	2	JCS837	utrophin - human
894	27	56.2	724	1	B32571	ribosomal protein	967	27	56.2	3433	1	S28381	hypothetical prote
895	27	56.2	726	2	T02409	probable Athila re	968	27	56.2	3972	2	S75251	ryanodine receptor
896	27	56.2	733	2	F84476	ribosomal protein	969	27	56.2	4859	2	S74173	giant protein p619
897	27	56.2	735	2	I51901	ribosomal protein	970	27	56.2	4861	2	S71752	ryanodine-binding
898	27	56.2	735	2	A53300	ribosomal protein	971	27	56.2	4868	2	B54161	ryanodine receptor
899	27	56.2	740	2	I38556	ribosomal protein	972	27	56.2	4869	2	S66572	ryanodine receptor
900	27	56.2	752	1	A32571	ATP-dependent RNA	973	27	56.2	4872	2	S27272	ryanodine receptor
901	27	56.2	759	2	G69258	DNA mismatch repai	974	27	56.2	4967	2	S72269	ryanodine receptor
902	27	56.2	776	2	E86830	hypothetical prote	975	27	56.2	4969	2	A37113	ryanodine receptor
903	27	56.2	776	2	T19900	secreted protease	976	27	56.2	5032	1	A35041	ryanodine receptor
904	27	56.2	781	2	G96991	protein H14A12.6 [	977	27	56.2	5035	1	I46646	ryanodine-binding
905	27	56.2	794	2	F88508		978	27	56.2	5037	1	A54161	

```
979 27 56.2 5037 2 B35041 ryanodine receptor
980 27 56.2 7463 2 T32248 CDA peptide synthe
981 27 56.2 13055 2 T16580 hypothetical prote
982 26.5 55.2 132 2 E86917 conserved hypothet
983 26.5 55.2 132 2 F70654 hypothetical prote
984 26.5 55.2 132 2 A72625 hypothetical prote
985 26.5 55.2 332 2 D95973 probable sugar upt
986 26.5 55.2 395 2 S78793 hypothetical prote
987 26.5 55.2 439 2 B82114 flagellum-specific
988 26.5 55.2 1186 2 S72229 meiotic recombinat
989 26.5 55.2 1205 2 T18517 procollagen N-endo
990 26.5 55.2 1292 2 T09227 galactose binding
991 26 54.2 36 2 A69827 cellular adhesion
992 26 54.2 56 2 A49619 hypothetical prote
993 26 54.2 64 2 D97979 hypothetical prote
994 26 54.2 64 2 G85659 hypothetical prote
995 26 54.2 77 2 B82441 hypothetical prote
996 26 54.2 77 2 D89887 hypothetical prote
997 26 54.2 84 2 S63410 hypothetical prote
998 26 54.2 84 2 S63323 hypothetical prote
999 26 54.2 84 2 D70967 hypothetical prote
1000 26 54.2 92 2 C27733 hypothetical prote

ALIGNMENTS

RESULT 1
S24712
Ig alpha chain - human
C:Species: Homo sapiens (man)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999
C:Accession: S24712
R:Tsapis, A.
submitted to the EMBL Data Library, August 1992
A:Reference number: S24708
A:Accession: S24712
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-83 <TSA>
A:Cross-references: EMBL:Z14963; NID:G28571; PIDN:CAA78686.1; PID:G28572
C:Keywords: immunoglobulin

Query Match 79.2%; Score 38; DB 2; Length 83;
Best Local Similarity 77.8%; Pred. No. 0.93;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9
DB 22 GPHETTTP 30

RESULT 2
T35421
probable regulatory protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
C:Accession: T35421
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21577
A:Accession: T35421
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-404 <OLI>
A:Cross-references: UNIPROT:Q9X7X6; EMBL:AL049485; PIDN:CAB39714.1; GSPDB:GN00070; SCOE
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC6A5.30C

Query Match 77.1%; Score 37; DB 2; Length 404;
Best Local Similarity 66.7%; Pred. No. 8.9;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

QY 1 GPHETITAL 9  
DB 296 GPEHVAAL 304

## RESULT 3

E86726  
cell division protein FtsY [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 09-Jul-2004  
A:Accession: E86726  
R:Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssi  
A:Reference number: A86625; MUID:21235186; PMID:11337471  
A:Accession: E86726  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-459 <STO>  
A:Cross-references: UNIPROT:Q9CHB9; GB:AE005176; PID:G12723735; PIDN:AAK04911.1; GSPDB:GN  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: ftsY  
C:Superfamily: cell division protein ftsY

Query Match 75.0%; Score 36; DB 2; Length 459;  
Best Local Similarity 87.5%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PHETITAL 9  
DB 374 PHETITAL 381

## RESULT 4

C88400  
protein H19M22.1 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
A:Accession: C88400  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_eleg  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: C88400  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1825 <STO>  
A:Cross-references: UNIPROT:Q9UB28; UNIPROT:Q9UB29; GB:chr\_III; PIDN:AAB94987.1; PID:G974  
C:Genetics:  
A:Gene: H19M22.1  
A:Map position: 3

Query Match 75.0%; Score 36; DB 2; Length 1825;  
Best Local Similarity 77.8%; Pred. No. 78;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
DB 288 GPHFTITGL 296

## RESULT 5

T32828  
hypothetical protein H19M22.1 - Caenorhabditis elegans (fragment)  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
A:Accession: T32828  
R:Wilson, R.; Wamsley, P.  
submitted to the EMBL Data Library, December 1997



A;Description: The sequence of C. elegans cosmid H19M22.

A;Reference number: Z21229

A;Accession: T32828

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1825 <MIL>

A;Cross-references: UNIPROT:Q9UB28; UNIPROT:Q9UB29; EMBL:AF040648; PIDN:AAB94997.1; GSPD

A;Experimental source: strain Bristol N2; clone H19M22

C;Genetics:

A;Gene: C8Sp.H19M22.1

A;Map position: 3

A;Introns: 271/2

Query Match 75.0%; Score 36; DB 2; Length 1825;

Best Local Similarity 77.8%; Pred. No. 78;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITL 9

|||||

288 GPHFTITGL 296

RESULT 6

A12205

hypothetical protein alr3200 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004

C;Accession: A12205

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: A12205

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-201 <KUR>

A;Cross-references: UNIPROT:Q8YS91; GB:BA000019; PIDN:BAB74899.1; PID:g17132295; GSPDB:G

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr3200

Query Match 72.9%; Score 35; DB 2; Length 201;

Best Local Similarity 71.4%; Pred. No. 11;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPHETIT 7

|||||

38 GPHETVS 44

RESULT 7

AB0936

N-acetyl-gamma-glutamyl-phosphate reductase [imported] - Salmonella enterica subsp. ente

C;Species: Salmonella enterica subsp. enterica serovar Typhi

A;Note: this species has also been called Salmonella typhi

C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C;Accession: AB0936

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A;Reference number: AB0502; MUID:21534947; PMID:11677608

A;Accession: AB0936

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-334 <PAR>

A;Cross-references: GB:AL513382; PIDN:CAD09508.1; PID:g16504625; GSPDB:GN00176

C;Genetics:

A;Gene: STV3752

C;Superfamily: N-acetyl-gamma-glutamyl-phosphate reductase

Query Match 72.9%; Score 35; DB 2; Length 334;

Best Local Similarity 87.5%; Pred. No. 19;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PHETITL 9

|||||

25 PHMTITL 32

RESULT 8

T06595

2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase (EC 2.7.6.3) - garden i

N;Contains: 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase (EC 2.7

C;Species: Pisum sativum (garden pea)

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004

C;Accession: T06595

R;Rebelle, F.; Macherel, D.; Mouillon, J.M.; Garin, J.; Douce, R.

EMBO J. 16, 947-957, 1997

A;Title: Folate biosynthesis in higher plants: purification and molecular cloning of a b

ochondria.

A;Reference number: Z15785; MUID:97224122; PMID:9118956

A;Accession: T06595

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-515 <REB>

A;Cross-references: UNIPROT:O04862; EMBL:Y08611; NID:g1934971; PIDN:CAA69903.1; PID:g1934

A;Note: parts of this sequence, including the amino end of the mature protein, were dete

C;Function: <HPPK>

A;Description: EC 2.7.6.3 [validated, MUID:97224122]

A;Pathway: folate biosynthesis

C;Function: <DHFS>

A;Description: EC 2.5.1.15 [validated, MUID:97224122]

A;Pathway: folate biosynthesis

A;Note: cofactor Mg(2+)

C;Superfamily: bifunctional folic acid synthesis protein; 2-amino-4-hydroxy-6-hydroxymeti

C;Keywords: diphosphotransferase; folate biosynthesis; mitochondrion

F;1-28/Domain: transit peptide (mitochondrion) #status predicted <TNP>

F;29-515/Product: bifunctional folic acid synthesis protein #status experimental <MAT>

F;48-179/Domain: 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase hom

F;232-488/Domain: dihydropterate synthase homology <DHS>

Query Match 72.9%; Score 35; DB 1; Length 515;

Best Local Similarity 66.7%; Pred. No. 31;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITL 9

|||||

107 GPHELLAAL 115

RESULT 9

B71404

hypothetical protein - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

A;Variety: Columbia

C;Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 09-Jul-2004

C;Accession: B71404

R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk

P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Giel

avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.

Nature 391, 485-488, 1998

A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech

erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans

C.; Chwalwis, N.

A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal

A;Reference number: A71400; MUID:98121113; PMID:9461215

A;Accession: B71404

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-724 <BBV>

A;Cross-references: UNIPROT:O23283; GB:Z97335; NID:g2244747; PID:e326878; PID:g2244781

C:Genetics:

A:Map position: 4COP9-4G3845

Query Match 72.9%; Score 35; DB 2; Length 724;

Best Local Similarity 66.7%; Pred. No. 45;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9

|||||:|

Db 526 GPHENIASL 534

RESULT 10

VHVVEV

structural polyprotein - eastern equine encephalomyelitis virus (strain 82V-2137)

N:Contains: 6K protein; coat protein C; membrane glycoprotein E1; membrane glycoprotein

C:Species: eastern equine encephalomyelitis virus

C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 09-Jul-2004

C:Accession: A26816

R:Chang, G.J.J.; Trent, D.W.

J. Gen. Virol. 68, 2129-2142, 1987

A:Title: Nucleotide sequence of the genome region encoding the 26S mRNA of eastern equin

A:Reference number: A26816; MUID:87282265; PMID:2886548

A:Accession: A26816

A:Molecule type: mRNA

A:Residues: 1-1239 <CHA>

A:Cross-references: UNIPROT:P08768; EMBL:X05816; NID:g62074; PIDN:CAA29261.1; PID:g62075

C:Superfamily: togavirus structural polyprotein

C:Keywords: coat protein; glycoprotein; transmembrane protein

F:1-259/Product: coat protein C #status predicted <CPC>

F:260-322/Product: membrane glycoprotein E3 #status predicted <MG3>

F:261-277/Domain: transmembrane #status predicted <TN1>

F:323-743/Product: membrane glycoprotein E2 #status predicted <MG2>

F:684-701/Domain: transmembrane #status predicted <TN2>

F:727-737/Domain: transmembrane #status predicted <TN3>

F:743-798/Product: 6K protein #status predicted <K6P>

F:777-798/Domain: transmembrane #status predicted <TN4>

F:799-1239/Product: membrane glycoprotein E1 #status predicted <MG1>

F:1211-1235/Domain: transmembrane #status predicted <TN5>

F:49,270,624,637,932/Binding site: carbohydrate (Asn) #status predicted

Query Match 72.9%; Score 35; DB 1; Length 1239;

Best Local Similarity 75.0%; Pred. No. 82;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PHETITAL 9

|||||:|

Db 289 PHETLTM 296

RESULT 11

VHVVEV

structural polyprotein - eastern equine encephalomyelitis virus (strain VA33[Ten Broeck])

N:Contains: 6K protein; coat protein C; membrane glycoprotein E1; membrane glycoprotein

C:Species: eastern equine encephalomyelitis virus

A:Note: host Equus caballus (domestic horse)

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004

C:Accession: A39992

R:Weaver, S.C.; Scott, T.W.; Rico-Hesse, R.

Virology 182, 774-784, 1991

A:Title: Molecular evolution of eastern equine encephalomyelitis virus in North America.

A:Reference number: A39992; MUID:91220727; PMID:2024496

A:Accession: A39992

A:Molecule type: Genomic RNA

A:Residues: 1-1240 <WEA>

A:Cross-references: UNIPROT:P27284; NID:g323696; PIDN:AAA42980.1; PID:g323697

A:Note: the authors translated the codon AGC for residue 836 as Arg and GUU for residue

C:Superfamily: togavirus structural polyprotein

C:Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein

F:1-260/Product: coat protein C #status predicted <CPC>

F:259-276/Domain: transmembrane #status predicted <TN1>

F:261-323/Product: membrane glycoprotein E3 #status predicted <EG3>

F:324-743/Product: membrane glycoprotein E2 #status predicted <EG2>

F:695-712/Domain: transmembrane #status predicted <TM2>

F:722-738/Domain: transmembrane #status predicted <TM3>

F:744-799/Product: 6K protein #status predicted <KP6>

F:781-799/Domain: transmembrane #status predicted <TM4>

F:800-1240/Product: membrane glycoprotein E1 #status predicted <EG1>

F:1212-1236/Domain: transmembrane #status predicted <TM5>

F:49,271,625,638,834,933/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 72.9%; Score 35; DB 1; Length 1240;

Best Local Similarity 75.0%; Pred. No. 82;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PHETITAL 9

|||||:|

Db 290 PHETLTM 297

RESULT 12

S26373

genome polyprotein - eastern equine encephalomyelitis virus

N:Contains: 6K protein; capsid protein C; envelope protein E1; envelope protein E2; envel

C:Species: eastern equine encephalomyelitis virus

C:Date: 06-Jan-1994 #sequence\_revision 01-Nov-1996 #text\_change 09-Jul-2004

C:Accession: S26373

R:Volchikov, V.E.; Volchkova, V.A.; Netesov, S.V.

Mol. Gen. Mikrobiol. Virusol. 5, 8-15, 1991

A:Title: Complete nucleotide sequence of the eastern equine encephalomyelitis virus genom

A:Reference number: S26369; MUID:91375524; PMID:1896061

A:Accession: S26373

A:Molecule type: mRNA

A:Residues: 1-1241 <VOL>

A:Cross-references: UNIPROT:Q66579; EMBL:X63135; NID:g59185; PIDN:CAA44845.1; PID:g59186

A:Note: sequence could not be checked because of bad print in paper

C:Superfamily: togavirus structural polyprotein

C:Keywords: capsid protein; envelope protein; glycoprotein; polyprotein

F:1-260/Product: capsid protein C #status predicted <CAP>

F:261-323/Product: envelope protein E3 #status predicted <EP3>

F:324-743/Product: envelope protein E2 #status predicted <EP2>

F:744-800/Product: 6K protein #status predicted <K6P>

F:801-1241/Product: envelope protein E1 #status predicted <EP1>

Query Match 72.9%; Score 35; DB 2; Length 1241;

Best Local Similarity 75.0%; Pred. No. 82;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PHETITAL 9

|||||:|

Db 290 PHETLTM 297

RESULT 13

A56605

structural polyprotein - eastern equine encephalomyelitis virus (strain 4789)

C:Species: eastern equine encephalomyelitis virus

C:Date: 11-Aug-1995 #sequence\_revision 11-Aug-1995 #text\_change 09-Jul-2004

C:Accession: A56605

R:Weaver, S.C.; Hegenbaugh, A.; Bellew, L.A.; Calisher, C.H.

Arch. Virol. 127, 305-314, 1992

A:Title: Genetic characterization of an antigenic subtype of eastern equine encephalomyel

A:Reference number: A56605; MUID:93090093; PMID:1280945

A:Status: preliminary

A:Molecule type: Genomic RNA

A:Residues: 1-1242 <WEA>

A:Cross-references: UNIPROT:Q08359; GB:L20951; NID:g405814; PIDN:AAA02897.1; PID:g305047

A:Note: sequence inconsistent with nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBI:119924, NCBI:P:119931)

C:Superfamily: togavirus structural polyprotein

C:Keywords: polyprotein

Query Match 72.9%; Score 35; DB 2; Length 1242;

Best Local Similarity 75.0%; Pred. No. 82;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PHETITAL 9  
||||:|  
Db 291 PHETLTM 298

RESULT 14  
S72350  
structural polyprotein - eastern equine encephalomyelitis virus  
N:Contains: GK protein; capsid protein; E1 protein; E2 protein; E3 protein  
C:Species: eastern equine encephalomyelitis virus  
C>Date: 04-May-1998 #sequence\_revision 15-May-1998 #text\_change 09-Jul-2004  
C:Accession: S72350  
C:Superfamily: Rv2626c  
R:Weaver, S.C.; Hagenbaugh, A.; Ballew, L.A.; Netesov, S.V.; Volchikov, V.E.; Chang, G.-J.  
Virology 197, 375-390, 1993  
A:Title: A comparison of the nucleotide sequences of eastern and western equine encephalomyelitis virus  
A:Reference number: S72349; MUID:94025587; PMID:8105605  
A:Accession: S72350  
A:Status: preliminary  
A:Molecule type: genomic RNA  
A:Residues: 1-1242 <WEA>  
A:Cross-references: UNIPROT:Q88790; EMBL:U01034; NID:g393006; PIDN:AAC53735.1; PID:g393006  
C:Superfamily: togavirus structural polyprotein

Query Match 72.9%; Score 35; DB 2; Length 1242;  
Best Local Similarity 75.0%; Pred. No. 82;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PHETITAL 9  
||||:|  
Db 291 PHETLTM 298

RESULT 15  
T14070  
peptide synthetase - Streptomyces fradiae (fragment)  
C:Species: Streptomyces fradiae  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: T14070  
R:Hosted, T.J.; Baltz, R.H.  
submitted to the EMBL Data Library, July 1997  
A:Reference number: Z17868  
A:Accession: T14070  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1324 <HOS>  
A:Cross-references: UNIPROT:O52048; EMBL:AF016696; NID:g2738764; PID:g2738765; PIDN:AACG  
C:Genetics:  
A:Note: cpe-1  
C:Superfamily: Mycobacterium tuberculosis mbe protein; acetate-CoA ligase homology; acyl carrier protein; phosphopantetheine; phosphoprotein  
F:188-842/Domain: acetate-CoA ligase homology <ACL>  
F:659-727/Domain: acyl carrier protein homology <ACP>  
F:691/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 72.9%; Score 35; DB 2; Length 1324;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPHETI 6  
|||||  
Db 659 GPHETI 664

RESULT 16  
A70573  
hypothetical protein Rv2626c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
C:Accession: A70573  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Cole, S.T.; Brosch, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Connor, R.; Davies, R.; Devlin, K.; Seeger, K.; Skelton, S.; Squares, S.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70573  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-143 <COL>  
A:Cross-references: UNIPROT:O06186; GB:Z95387; GB:AL123456; NID:g3261763; PIDN:CAB08616.1  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv2626c  
C:Superfamily: conserved hypothetical protein yhcV; CBS homology  
F:77-124/Domain: CBS homology <CBS>

Query Match 70.8%; Score 34; DB 2; Length 143;  
Best Local Similarity 75.0%; Pred. No. 12;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPHETITA 8  
||||:|  
Db 16 GEHETITA 23

RESULT 17  
T40860  
probable alpha-amylase precursor SPC11E10.09c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T40860; T41181  
R:Ramsperger, U.; Pohl, T.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, October 1999  
A:Reference number: Z21952  
A:Accession: T40860  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-478 <RAM>  
A:Cross-references: UNIPROT:Q10427; EMBL:AL121783; NID:g6016988; PIDN:CAB57851.1; PID:g6016988  
A:Experimental source: strain 972h-; cosmid c11E10  
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Harris, D.  
submitted to the EMBL Data Library, April 1999  
A:Reference number: Z21976  
A:Accession: T41181  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-203 <LYN>  
A:Cross-references: EMBL:AL049662; NID:g4678680; PIDN:CAB41221.1; PID:g4678681; GSPDB:G6N  
A:Experimental source: strain 972h-; cosmid c188  
C:Genetics:  
A:Gene: SPDB:SPC11E10.09c; SPDB:SPC1188.01c  
A:Map position: 3  
A:Introns: 320/3; 468/3  
C:Superfamily: Aspergillus alpha-amylase; alpha-amylase core homology

Query Match 70.8%; Score 34; DB 2; Length 478;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PHETIT 7  
|||||  
Db 168 PHETIT 173

RESULT 18  
G82971  
probable ferredoxin PAS399 [imported] - Pseudomonas aeruginosa (strain PAO1)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
C:Accession: G82971  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Lim, adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, ; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000



C:Species: *Bacillus subtilis*  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
 C:Accession: D69675; I40024  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertex  
 C.; Bron, S.; Brouillet, S.; Bruchi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen  
 A:Authors: Koningstein, G.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Magada, S.; Mausel  
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seron  
 Akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: D69675  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-511 <UN>  
 A:Cross-references: UNIPROT:P39773; GB:Z99121; GB:AL009126; NID:G2635827; PIDN:CAB15396.  
 A:Experimental source: strain 168  
 R:Leyva-Vazquez, M.A.; Setlow, P.  
 J. Bacteriol. 176, 3903-3910, 1994  
 A:Title: Cloning and nucleotide sequences of the genes encoding triose phosphate isomera  
 A:Reference number: I40022; MUID:94292408; PMID:8021172  
 A:Accession: I40024  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-232, 'C', 234-431, 'MV', 434-444, 'D', 446-511 <RES>  
 A:Cross-references: GB:I29475; NID:G460256; PIDN:AAA21680.1; PID:G460258  
 C:Genetics:  
 A:Gene: pgm  
 C:Function:  
 A:Description: EC 5.4.2.1 [validated, MUID:94292408]  
 A:Note: 2,3-bisphosphoglycerate-independent but Mn2+-dependent enzyme [validated, MUID:9  
 C:Superfamily: cofactor-independent phosphoglycerate mutase  
 C:Keywords: intramolecular transferase; isomerase  
 Query Match 68.8%; Score 33; DB 2; Length 511;  
 Best Local Similarity 71.4%; Pred. No. 80;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 PHETITA 8  
 DB 41 PHOTLTA 47  
 RESULT 24  
 E86106  
 yidB protein [similarity] - *Escherichia coli* (strain O157:H7, substrain EDL933)  
 C:Species: *Escherichia coli*  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
 C:Accession: E86106  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: E86106  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-557 <STO>  
 A:Cross-references: UNIPROT:O8XEJ3; GB:AB005174; NID:G12519084; PIDN:AAG59313.1; GSPDB:G  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Superfamily: *Escherichia coli* yidB protein  
 Query Match 68.8%; Score 33; DB 2; Length 557;

Best Local Similarity 62.5%; Pred. No. 88;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 PHETITAL 9  
 DB 347 PHQNTAL 354  
 RESULT 25  
 S56342  
 yidB protein - *Escherichia coli* (strain K-12)  
 C:Species: *Escherichia coli*  
 C:Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 09-Jul-2004  
 R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
 Nucleic Acids Res. 23, 2105-2119, 1995  
 A:Title: Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.4  
 A:Reference number: S56314; MUID:95334362; PMID:7610040  
 A:Accession: S56342  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-557 <BUR>  
 A:Cross-references: UNIPROT:O8XEJ3; EMBL:U14003; NID:G1263172; PIDN:AAA97013.1; PID:G5636.  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of *Escherichia coli* K-12.  
 A:Reference number: A64720; MUID:97426617; PMID:9278503  
 A:Accession: A65221  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-557 <BLAT>  
 A:Cross-references: GB:AE000483; GB:U00096; NID:G2367351; PIDN:AAC77075.1; PID:G1790553;  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:  
 A:Gene: yidB  
 C:Superfamily: *Escherichia coli* yidB protein  
 Query Match 68.8%; Score 33; DB 2; Length 557;  
 Best Local Similarity 62.5%; Pred. No. 88;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 PHETITAL 9  
 DB 347 PHQNTAL 354  
 RESULT 26  
 H91265  
 hypothetical protein ECs5096 [imported] - *Escherichia coli* (strain O157:H7, substrain RIN  
 C:Species: *Escherichia coli*  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
 C:Accession: H91265  
 R:Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 gawawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genom  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: H91265  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-557 <HAY>  
 A:Cross-references: UNIPROT:O8XEJ3; GB:BA000007; PIDN:BA838519.1; PID:G13364573; GSPDB:G  
 A:Experimental source: strain O157:H7, substrain RIMD 0509952  
 C:Genetics:  
 A:Gene: ECs5096  
 C:Superfamily: *Escherichia coli* yidB protein  
 Query Match 68.8%; Score 33; DB 2; Length 557;  
 Best Local Similarity 62.5%; Pred. No. 88;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

QY      2 PHEITITAL 9
      ||: :|||
Db      347 PHQNTVAL 354

RESULT 27
D87380
hypothetical protein CC1056 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C:Accession: D87380
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: AB7249; MUID:21173698; PMID:11259647
A:Accession: D87380
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-613 <STO>
A:Cross-references: UNIPROT:Q9A9D2; GB:AE005673; NID:gl3422354; PIDN:AAK23040.1; GSPDB:G
C:Genetics:
A:Gene: CC1056

      Query Match      68.8%; Score 33; DB 2; Length 613;
      Best Local Similarity 66.7%; Pred. No. 99;
      Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 GPHETITAL 9
      ||: :|||
Db      44 GPSETLTGL 52

RESULT 28
S49788
probable membrane protein Y1L092w - Yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein Y1910.04
C:Species: Saccharomyces cerevisiae
C>Date: 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 09-Jul-2004
C:Accession: S49788
R:Connor, R.; Churcher, C.
submitted to the EMBL Data Library, November 1994
A:Reference number: S49786
A:Accession: S49788
A:Molecule type: DNA
A:Residues: 1-633 <CON>
A:Cross-references: UNIPROT:P40497; GB:Z47047; EMBL:Z46728; NID:g603997; PID:g763254; GS
C:Genetics:
A:Gene: MIPS:Y1L092w
A:Cross-references: SGD:S0001354
A:Map position: 9L
C:Superfamily: Saccharomyces cerevisiae probable membrane protein Y1L092w
C:Keywords: transmembrane protein
F:99-115/Domain: transmembrane #status predicted <TM>

      Query Match      68.8%; Score 33; DB 2; Length 633;
      Best Local Similarity 85.7%; Pred. No. 1e+02;
      Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 PHEITITA 8
      |||||
Db      566 PHETINA 572

RESULT 29
WNB6W6
capsid protein - human herpesvirus 1 (strain 17)
C:Species: human herpesvirus 1
C>Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 09-Jul-2004
C:Accession: H30084
R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perz
J. Gen. Virol. 69, 1531-1574, 1988

```

```

A:Title: The complete DNA sequence of the long unique region in the genome of herpes sim
A:Reference number: A30083; MUID:88274327; PMID:2839594
A:Accession: H30084
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-635 <MCG>
A:Cross-references: UNIPROT:P10210; GB:X14112; NID:gl944536; PIDN:CAA32318.1; PID:g59526;
C:Genetics:
A:Gene: UL26
C:Superfamily: varicella-zoster virus gene 33 protein
C:Superfamily: capsid protein

      Query Match      68.8%; Score 33; DB 1; Length 635;
      Best Local Similarity 77.8%; Pred. No. 1e+02;
      Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 GPHETITAL 9
      ||: :|||
Db      473 GTNETITAL 481

RESULT 30
D82452
anaerobic ribonucleoside-triphosphate reductase VCA0511 [imported] - Vibrio cholerae [str
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: D82452
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833; PMID:10952301
A:Accession: D82452
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-706 <HEI>
A:Cross-references: UNIPROT:Q9KM77; GB:AE004381; GB:AE003853; NID:g9657902; PIDN:AAF96414
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0511
A:Map position: 2
C:Superfamily: Escherichia coli oxygen-sensitive ribonucleoside-triphosphate reductase; c
ide-triphosphate reductase middle homology; rubredoxin homology
C:Keywords: iron; metalloprotein
F:644,647,662,665/Binding site: iron (Cys) #status predicted

      Query Match      68.8%; Score 33; DB 2; Length 706;
      Best Local Similarity 77.8%; Pred. No. 1.2e+02;
      Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GPHETITAL 9
      |||||
Db      492 GVHETITAL 500

RESULT 31
A64047
ribonucleoside-triphosphate reductase, oxygen-sensitive (EC 1.17.4.-) - Haemophilus infl
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 19-Jan-1996 #text_change 09-Jul-2004
C:Accession: A64047
R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: A64047
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-707 <TIGR>

```

A;Cross-references: UNIPROT:P43752; GB:U32693; NID:G1573021; PIDN:AA21751.1;  
C;Comment: This enzyme must be activated by the anaerobic ribonucleotide reductase activ-  
es peptide cleavage.  
C;Function:  
A;Description: catalyzes the reduction by reduced thioredoxin of a ribonucleoside tripho-  
A;Pathway: deoxyribonucleotide biosynthesis  
A;Note: this enzyme is different both from ribonucleoside-triphosphate reductase (EC 1.1.1  
nds iron and has a tyrosyl radical  
C;Superfamily: Escherichia coli oxygen-sensitive ribonucleoside-triphosphate reductase;  
ide-triphosphate reductase middle homology; rubredoxin homology  
F;96-279/Domain: oxygen-sensitive ribonucleoside-triphosphate reductase carboxyl-termin-  
F;305-706/Domain: oxygen-sensitive ribonucleoside-triphosphate reductase middle homology  
F;641-671/Domain: rubredoxin homology #status atypical <RUB>  
F;645,648,663,666/Binding site: iron (Cys) #status predicted  
F;662/Active site: Gly (stable glycyl radical) #status predicted

Query Match 68.8%; Score 33; DB 1; Length 707;  
Best Local Similarity 77.8%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
| | | | | | |  
Db 494 GIHETINAL 502

RESULT 32

A47331  
ribonucleoside-triphosphate reductase, oxygen-sensitive (EC 1.17.4.-) - Escherichia coli  
N;Alternate names: anaerobic ribonucleotide reductase  
C;Species: Escherichia coli  
C;Date: 21-Sep-1993 #sequence revision 19-Jan-1996 #text\_change 09-Jul-2004  
C;Accession: A47331; S56464; J65236  
R;Sun, X.; Harder, J.; Krook, M.; Jorvall, H.; Sjoberg, B.M.; Reichard, P.  
Proc. Natl. Acad. Sci. U.S.A. 90, 577-581, 1993  
A;Title: A possible glycine radical in anaerobic ribonucleotide reductase from Escherich  
A;Reference number: A47331; MUID:93133831; PMID:8421692  
A;Accession: A47331  
A;Molecule type: DNA  
A;Residues: 1-256; R', 421-712 <SUN>  
A;Cross-references: UNIPROT:P28903; GB:L06097; NID:G146968; PIDN:AA24226.1; PID:G146970  
A;Note: sequence extracted from NCBI backbone (NCBIN:122818, NCBI:P122819)  
A;Note: parts of this sequence, including the amino end of the mature protein, were dete-  
R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
Nucleic Acids Res. 23, 2105-2119, 1995  
A;Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.  
A;Reference number: S56314; MUID:95334362; PMID:7610040  
A;Accession: S56464  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-712 <BUR>  
A;Cross-references: EMBL:U14003; NID:G1263172; PIDN:AAA97135.1; PID:G537080  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co-  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503  
A;Accession: A65236  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-712 <BLAT>  
A;Cross-references: GB:AE000495; NID:G2367361; PIDN:AACT77195.1; PID:G1790686;  
A;Experimental source: strain K-12, substrain MG1655  
C;Comment: This enzyme must be activated by the anaerobic ribonucleotide reductase activ-  
es peptide cleavage.  
C;Genetics:  
A;Gene: nrdd  
A;Map position: 96 min  
C;Complex: homodimer  
C;Function:  
A;Description: catalyzes the reduction by reduced thioredoxin of a ribonucleoside tripho-  
A;Pathway: deoxyribonucleotide biosynthesis

A;Note: this enzyme is different both from ribonucleoside-triphosphate reductase (EC 1.1.1;  
nds iron and has a tyrosyl radical  
C;Superfamily: Escherichia coli oxygen-sensitive ribonucleoside-triphosphate reductase;  
ide-triphosphate reductase middle homology; rubredoxin homology  
C;Keywords: deoxyribonucleotide biosynthesis; homodimer; iron; metalloprotein; oxidoreduc-  
F;96-279/Domain: oxygen-sensitive ribonucleoside-triphosphate reductase carboxyl-termin-  
F;305-706/Domain: oxygen-sensitive ribonucleoside-triphosphate reductase middle homology  
F;641-671/Domain: rubredoxin homology #status atypical <RUB>  
F;644,647,662,665/Binding site: iron (Cys) #status predicted  
F;681/Active site: Gly (stable glycyl radical) #status predicted

Query Match 68.8%; Score 33; DB 1; Length 712;  
Best Local Similarity 77.8%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
| | | | | | |  
Db 493 GIHETINAL 501

RESULT 33

AC1058  
ribonucleoside-triphosphate reductase (EC 1.17.4.2) - Salmonella enterica subsp. enterica  
C;Species: Salmonella enterica subsp. enterica serovar Typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AC1058  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AC1058  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-712 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD06912.1; PID:G16505560; GSPDB:GN00176  
C;Genetics:  
A;Gene: nrdd  
C;Superfamily: Escherichia coli oxygen-sensitive ribonucleoside-triphosphate reductase;  
ide-triphosphate reductase middle homology; rubredoxin homology  
C;Keywords: iron; metalloprotein; oxidoreductase  
F;644,647,662,665/Binding site: iron (Cys) #status predicted

Query Match 68.8%; Score 33; DB 2; Length 712;  
Best Local Similarity 77.8%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
| | | | | | |  
Db 493 GIHETINAL 501

RESULT 34

G91280  
anaerobic ribonucleoside-triphosphate reductase [imported] - Escherichia coli (strain O1  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C;Accession: G91280  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno-  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: G91280  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-712 <HAY>  
A;Cross-references: UNIPROT:Q8XCE2; GB:BA000007; PIDN:BA38638.1; PID:G13364692; GSPDB:G  
A;Experimental source: strain O157:H7, substrain RIMD 0509952  
C;Genetics:

A:Gene: EC5215  
C:Superfamily: Escherichia coli oxygen-sensitive ribonucleoside-triphosphate reductase;  
ide-triphosphate reductase middle homology; rubredoxin homology  
C:Keywords: iron; metalloprotein  
F:644,647,662,665/Binding site: iron (Cys) #status predicted

Query Match 68.8%; Score 33; DB 2; Length 712;  
Best Local Similarity 77.8%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
| | | | | | |  
DB 493 GIHETINAL 501

RESULT 35  
G86121  
A:Gene: anaerobic ribonucleoside-triphosphate reductase [imported] - Escherichia coli (strain O1  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: G86121  
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: G86121  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-712 <STO>  
A:Cross-references: UNIPROT:O8XCE2; GB:AE005174; NID:gl12519242; PIDN:AG59435.1; GSPDB:G  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: nrdd  
C:Superfamily: Escherichia coli oxygen-sensitive ribonucleoside-triphosphate reductase;  
ide-triphosphate reductase middle homology; rubredoxin homology  
C:Keywords: iron; metalloprotein  
F:644,647,662,665/Binding site: iron (Cys) #status predicted

Query Match 68.8%; Score 33; DB 2; Length 712;  
Best Local Similarity 77.8%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
| | | | | | |  
DB 493 GIHETINAL 501

RESULT 36  
AG0419  
A:Gene: ribonucleoside-triphosphate reductase (EC 1.17.4.2) [imported] - Yersinia pestis (strain  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AG0419  
R:Parkhill, J.; Wren, B.W.; Thomeon, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.  
do-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AG0419  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-712 <KUR>  
A:Cross-references: UNIPROT:Q8ZBG2; GB:AL590842; PIDN:CAC92683.1; PID:g15981378; GSPDB:G  
C:Genetics:  
A:Gene: nrdd  
C:Superfamily: Escherichia coli oxygen-sensitive ribonucleoside-triphosphate reductase;  
ide-triphosphate reductase middle homology; rubredoxin homology  
C:Keywords: iron; metalloprotein; oxidoreductase  
F:644,647,662,665/Binding site: iron (Cys) #status predicted

Query Match 68.8%; Score 33; DB 2; Length 712;

Best Local Similarity 77.8%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
| | | | | | |  
DB 492 GIHETINAL 500

RESULT 37  
S54595  
A:Gene: probable membrane protein YMR288w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein YMR021.14  
C:Species: Saccharomyces cerevisiae  
C:Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004  
C:Accession: S54595  
R:Pearson, D.; Bowman, S.  
submitted to the EMBL Data Library, May 1995  
A:Reference number: S54582  
A:Accession: S54595  
A:Molecule type: DNA  
A:Residues: 1-971 <PEA>  
A:Cross-references: UNIPROT:P49955; EMBL:Z49704; NID:g8255540; PID:g8255554; GSPDB:GN00013;  
A:Experimental source: strain AB972  
C:Genetics:  
A:Gene: SGD:HS155; MIPS:YMR288w  
A:Cross-references: SGD:S0004901  
A:Map position: 13R  
C:Keywords: transmembrane protein  
F:333-349/Domain: transmembrane #status predicted <TM1>  
F:777-793/Domain: transmembrane #status predicted <TM2>

Query Match 68.8%; Score 33; DB 2; Length 971;  
Best Local Similarity 55.6%; Pred. No. 1.6e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
| | | | | | |  
DB 758 GPHDVLVAL 766

RESULT 38  
T30544  
A:Gene: major surface glycoprotein - Pneumocystis carinii (fragment)  
C:Species: Pneumocystis carinii  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 15-Jun-2001  
C:Accession: T30544  
R:Mei, Q.; Turner, R.E.; Sorial, V.; Klivington, D.; Angus, C.W.; Kovacs, J.A.  
Infect. Immun. 66, 4268-4273, 1998  
A:Title: Characterization of major surface glycoprotein genes of human Pneumocystis carii  
A:Reference number: Z17905; MUID:98380374; PMID:9712777  
A:Accession: T30544  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1008 <MEI>  
A:Cross-references: EMBL:AF033212; NID:g3560520; PID:g3560521; PIDN:AAC34975.1  
C:Genetics:  
A:Gene: MSG  
C:Superfamily: Pneumocystis carinii major surface glycoprotein MSG100

Query Match 68.8%; Score 33; DB 2; Length 1008;  
Best Local Similarity 83.3%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PHETIT 7  
| | | | | | |  
DB 920 PHETITV 925

RESULT 39  
T30542  
A:Gene: major surface glycoprotein - Pneumocystis carinii (fragment)  
C:Species: Pneumocystis carinii  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 15-Jun-2001



C;Accession: T30542  
R;Mei, Q.; Turner, R.E.; Sorial, V.; Klivington, D.; Angus, C.W.; Kovacs, J.A.  
Infect. Immun. 66, 4268-4273, 1998  
A;Title: Characterization of major surface glycoprotein genes of human Pneumocystis carinii  
A;Reference number: Z17905; MUID:98380374; PMID:9712777  
A;Accession: T30542  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1017 <MEI>  
A;Cross-references: EMBL:AF033210; NID:G3560516; PID:G3560517; PIDN:AAC34973.1  
C;Genetics:  
A;Gene: MSG  
C;Superfamily: Pneumocystis carinii major surface glycoprotein MSG100

Query Match 68.8%; Score 33; DB 2; Length 1017;  
Best Local Similarity 83.3%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PHETIT 7  
Db 912 PHETVT 917  
|||||

RESULT 40  
T30543  
major surface glycoprotein - Pneumocystis carinii (fragment)  
C;Species: Pneumocystis carinii  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 15-Jun-2001  
C;Accession: T30543  
R;Mei, Q.; Turner, R.E.; Sorial, V.; Klivington, D.; Angus, C.W.; Kovacs, J.A.  
Infect. Immun. 66, 4268-4273, 1998  
A;Title: Characterization of major surface glycoprotein genes of human Pneumocystis carinii  
A;Reference number: Z17905; MUID:98380374; PMID:9712777  
A;Accession: T30543  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1022 <MEI>  
A;Cross-references: EMBL:AF033211; NID:G3560518; PID:G3560519; PIDN:AAC34974.1  
C;Genetics:  
A;Gene: MSG  
C;Superfamily: Pneumocystis carinii major surface glycoprotein MSG100

Query Match 68.8%; Score 33; DB 2; Length 1022;  
Best Local Similarity 83.3%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PHETIT 7  
Db 915 PHETVT 920  
|||||

RESULT 41  
C36792  
hypothetical protein ORF56 - ictalurid herpesvirus 1 (strain auburn 1)  
C;Species: ictalurid herpesvirus 1  
A;Note: host Ictalurus punctatus (channel catfish)  
C;Date: 17-Aug-1992 #sequence\_revision 17-Aug-1992 #text\_change 09-Jul-2004  
C;Accession: C36792  
R;Davison, A.J.  
submitted to GenBank, January 1992  
A;Description: Channel catfish virus: a new type of herpesvirus.  
A;Reference number: A36804  
A;Accession: C36792  
A;Molecule type: DNA  
A;Residues: 1-1179 <DAV>  
A;Cross-references: UNIPROT:Q00099; GB:M75136; NID:G331209; PIDN:AAA8159.1; PID:G331266  
R;Davison, A.J.  
Virology 186, 9-14, 1992  
A;Title: Channel catfish virus: a new type of herpesvirus.  
A;Reference number: A39447; MUID:92087490; PMID:1727613  
A;Contents: annotation  
A;Note: neither protein nor nucleic acid sequence is given  
C;Genetics:

A;Gene: 56  
C;Superfamily: ictalurid herpesvirus 1 hypothetical protein ORF56

Query Match 68.8%; Score 33; DB 2; Length 1179;  
Best Local Similarity 85.7%; Pred. No. 2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPHEITIT 7  
Db 828 GPHGTIT 834  
|||||

RESULT 42  
MWXR31  
lambda 3 protein - reovirus type 1 (strain Lang)  
N;Alternate names: minor core protein  
C;Species: reovirus type 1  
C;Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 09-Jul-2004  
C;Accession: A30121  
R;Wiener, J.R.; Joklik, W.K.  
Virology 169, 194-203, 1989  
A;Title: The sequences of the reovirus serotype 1, 2, and 3 L1 genome segments and analysis  
A;Reference number: A94390; MUID:89163254; PMID:2922925  
A;Accession: A30121  
A;Molecule type: genomic RNA  
A;Residues: 1-1267 <WIE>  
A;Cross-references: UNIPROT:P17376; GB:M24734; NID:G499863  
A;Note: This sequence, which matches the sequence attributed to type 1 in Fig. 2, matches the translations in entries REO3LAM3P and REO3LAM3P now differ only by the sequence correction  
C;Comment: See also PIR:MWXR33.  
C;Genetics:  
A;Map position: segment L1  
C;Superfamily: reovirus lambda 3 protein  
C;Keywords: core protein

Query Match 68.8%; Score 33; DB 1; Length 1267;  
Best Local Similarity 71.4%; Pred. No. 2.2e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 PHETITA 8  
Db 105 PHETLTS 111  
|||||

RESULT 43  
MWXR32  
lambda 3 protein - reovirus type 2 (strain D5/Jones)  
N;Alternate names: minor core protein  
C;Species: reovirus type 2  
A;Note: host Homo sapiens (man)  
C;Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 09-Jul-2004  
C;Accession: B30121  
R;Wiener, J.R.; Joklik, W.K.  
Virology 169, 194-203, 1989  
A;Title: The sequences of the reovirus serotype 1, 2, and 3 L1 genome segments and analysis  
A;Reference number: A94390; MUID:89163254; PMID:2922925  
A;Accession: B30121  
A;Molecule type: genomic RNA  
A;Residues: 1-1267 <WIE>  
A;Cross-references: UNIPROT:P17377; GB:M31057; NID:G499865; PIDN:AAA47245.1; PID:G499866  
C;Genetics:  
A;Map position: segment L1  
C;Superfamily: reovirus lambda 3 protein  
C;Keywords: core protein

Query Match 68.8%; Score 33; DB 1; Length 1267;  
Best Local Similarity 71.4%; Pred. No. 2.2e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 PHETITA 8  
Db 105 PHETLTS 111  
|||||

## RESULT 44

MXR33  
lambda 3 protein - reovirus type 3 (strain Dearing)  
N:Alternate names: minor core protein  
C:Species: reovirus type 3

C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 09-Jul-2004  
C:Accession: C30121  
R:Wiener, J.R.; Joklik, W.K.

Virology 169, 194-203, 1989  
A:Title: The sequences of the reovirus serotype 1, 2, and 3 L1 genome segments and analyses REO11AM3P and REO31AM3P now differ only by the sequence correction apparently made  
A:Reference number: A94390; MUID:89163254; PMID:2922925  
A:Accession: C30121

A:Molecule type: genomic RNA

A:Residues: 1-1267 <WIR>

A:CROSS-references: UNIPROT:P17378; GB:M31058; NID:G499867; GB:M24734; NID:G499863; PIDN: A:Note: this sequence, which matches the sequence attributed to type 3 in Fig. 2, matches ntries REO11AM3P and REO31AM3P  
C:Comment: See also PIR:MXR31.

C:Genetics:

A:Map position: segment L1

C:Superfamily: reovirus lambda 3 protein

C:Keywords: core protein

Query Match 68.8%; Score 33; DB 1; Length 1267;  
Best Local Similarity 71.4%; Pred. No. 2.2e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 PHETITA 8

DB 105 PHETILTS 111

## RESULT 45

T38353  
serine hydroxymethyltransferase - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T38353

R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, August 1995

A:Reference number: Z21787

A:Accession: T38353

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-467 <MUR>

A:CROSS-references: UNIPROT:Q13972; EMBL:Z98601; PIDN:CAB11269.1; GSPDB:GN00066; SPDB:SF  
A:Experimental source: strain 972h-; cosmid C24C9

C:Genetics:

A:Gene: SPDB:SPAC24C9.12c

A:Map position: 1

C:Superfamily: serine/glycine hydroxymethyltransferase

Query Match 67.7%; Score 32.5; DB 2; Length 467;  
Best Local Similarity 80.0%; Pred. No. 93;  
Matches 8; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 GPH-ETITAL 9

DB 288 GPHNHTITAL 297

## RESULT 46

T37918

serine hydroxymethyltransferase - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T37918

R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.

submitted to the EMBL Data Library, December 1995

A:Reference number: Z21754

A:Accession: T37918

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-472 <CON>

A:CROSS-references: UNIPROT:Q10104; EMBL:Z68198; PIDN:CAA92384.1; GSPDB:GN00066; SPDB:SP;  
A:Experimental source: strain 972h-; cosmid c18G6

C:Genetics:

A:Gene: SPDB:SPAC18G6.04c

A:Map position: 1

C:Superfamily: serine/glycine hydroxymethyltransferase

Query Match 67.7%; Score 32.5; DB 2; Length 472;  
Best Local Similarity 80.0%; Pred. No. 94;  
Matches 8; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 GPH-ETITAL 9

DB 294 GPHNHTITAL 303

## RESULT 47

A42241

glycine hydroxymethyltransferase (EC 2.1.1.2.1), cytosolic - Neurospora crassa  
N:Alternate names: serine hydroxymethyltransferase  
C:Species: Neurospora crassa

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004

C:Accession: A42241

R:McClung, C.R.; Davis, C.R.; Page, K.M.; Denome, S.A.

Mol. Cell. Biol. 12, 1412-1421, 1992

A:Title: Characterization of the format (for) locus, which encodes the cytosolic serine  
A:Reference number: A42241; MUID:92195285; PMID:1532227

A:Accession: A42241

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-479 <MCC>

A:CROSS-references: UNIPROT:P34898; GB:M81918

C:Superfamily: serine/glycine hydroxymethyltransferase

C:Keywords: cytosol; phosphoprotein; pyridoxal phosphate; transferase

F:249/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 67.7%; Score 32.5; DB 1; Length 479;  
Best Local Similarity 80.0%; Pred. No. 95;  
Matches 8; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 GPH-ETITAL 9

DB 294 GPHNHTITAL 303

## RESULT 48

T03782

probable lipid transfer protein - rice

C:Species: Oryza sativa (rice)

C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004

C:Accession: T03782

R:Vignols, F.; Lund, G.; Pham, S.; Tremousaygue, D.; Grelllet, F.; Kader, J.; Puidomenech

Gene 142, 265-270, 1994

A:Title: Characterization of a rice gene coding for a lipid transfer protein.

A:Reference number: Z15081; MUID:94252578; PMID:8194762

A:Accession: T03782

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-117 <VIG>

A:CROSS-references: UNIPROT:Q42999; EMBL:Z23271; NID:G510336; PIDN:CAA80809.1; PID:G51033

A:Experimental source: cv. IR36

C:Genetics:

A:Introns: 115/3

C:Superfamily: phospholipid transfer protein

Query Match 66.7%; Score 32; DB 2; Length 117;  
Best Local Similarity 55.6%; Pred. No. 25;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9

DB 111 111 111

Db 20 GPHTTMAAI 28

RESULT 49

T02042

lipid transfer protein LPT II - rice

C;Species: Oryza sativa (rice)

C;Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 09-Jul-2004

C;Accession: T02042

R;Lee, M.C.; Kim, C.S.; Eun, M.Y.

submitted to the EMBL Data Library, August 1997

A;Description: Isolation and characterization of lipid transfer protein from rice.

A;Reference number: Z14508

A;Accession: T02042

A;Status: preliminary; translated from GE/EMBL/DBDJ

A;Molecule type: mRNA

A;Residues: 1-118 <LEE>

A;Cross-references: UNIPROT:O22483; EMBL:AF017359; NID:g2407272; PIDN:AAB70539.1; PID:g2

A;Experimental source: strain Milyang 23

C;Superfamily: phospholipid transfer protein

Query Match 66.7%; Score 32; DB 2; Length 118;  
Best Local Similarity 55.6%; Pred.No. 25;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHTTITAL 9

||| | | |

Db 20 GPHTTMAAI 28

RESULT 50

T03300

probable lipid transfer protein precursor - rice

C;Species: Oryza sativa (rice)

C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004

C;Accession: T03300

R;Li, G.

submitted to the EMBL Data Library, July 1995

A;Reference number: Z14895

A;Accession: T03300

A;Status: preliminary; translated from GE/EMBL/DBDJ

A;Molecule type: mRNA

A;Residues: 1-118 <LIA>

A;Cross-references: UNIPROT:Q42978

A;Experimental source: strain Guang-lu-ai 4, shoot

C;Genetics:

A;Gene: Ltp2

C;Superfamily: phospholipid transfer protein

F;1-26/Domain: signal sequence #status predicted <SIG>

F;27-118/Product: lipid transfer protein #status predicted <MAT>

Query Match 66.7%; Score 32; DB 2; Length 118;  
Best Local Similarity 55.6%; Pred.No. 25;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHTTITAL 9

||| | | |

Db 20 GPHTTMAAI 28

Search completed: July 27, 2005, 12:42:07

Job time : 56 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2005, 12:03:22 ; Search time 40 Seconds  
(without alignments)  
538.814 Million cell updates/sec

Title: US-10-623-429-9\_COPY\_404\_627  
Perfect score: 1237  
Sequence: 1 AIAADROAGGLPAAAGDHGI.....ALVNASSAAHVNVDTPARAAD 224

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	670.5	54.2	635	1 WMBEW6	capsid protein - h
2	221.5	17.9	522	2 S52216	viral proteinase -
3	215	17.4	699	2 T05225	extensin homolog F
4	211	17.1	646	1 W2BEC8	68.6K capsid prote
5	206	16.7	727	2 C84534	hypothetical prote
6	204	16.5	620	2 S06733	hydroxyproline-ric
7	200	16.2	464	2 S22697	extensin - Volvox
8	198	16.0	760	2 T08291	extensin homolog T
9	196	15.8	929	2 C96623	hypothetical prote
10	195.5	15.8	599	2 T10798	pherophorin-S - Vo
11	195	15.8	134	2 JC5572	proline-rich prote
12	194.5	15.7	485	2 A33647	sulfated surface g
13	193	15.6	847	2 F96531	hypothetical prote
14	191.5	15.5	217	2 T09965	extensin CYC17 pre
15	190.5	15.4	551	2 S57447	HPBRII-7 protein -
16	190.5	15.4	647	2 T42579	capsid protein 35
17	190	15.4	464	2 A47655	spliceosome-associ
18	190	15.4	1188	2 S49915	extensin-like prot
19	188.5	15.2	428	2 E71415	probable coll wall
20	187.5	15.2	839	2 T04859	extensin homolog F
21	187	15.1	368	2 C29356	hydroxyproline-ric
22	187	15.1	907	2 E96636	hypothetical prote
23	186.5	15.1	907	2 T28872	hypothetical prote
24	186	15.0	1006	2 G86292	hypothetical prote
25	185	15.0	242	2 S54156	extensin-like prot
26	185	15.0	416	2 T34279	hypothetical prote
27	185	15.0	760	2 F86387	probable pro kinas
28	184.5	14.9	132	2 S14970	extensin class I (
29	184.5	14.9	1018	2 T43168	hypothetical prote

hypothetical prote  
probable proline-r  
hypothetical proce  
immediate-early pr  
extensin-like cell  
hypothetical prote  
C-terminal domain-  
extensin-like prot  
extensin precursor  
pistil extensin-li  
chitinase (EC 3.2.  
unknown protein [i  
hydroxyproline-ric  
nodulin precursor  
hydroxyproline-ric  
Wiskott-Aldrich sy

30 184.5 14.9 1259 2 T16038  
31 183 14.8 191 2 F84522  
32 183 14.8 222 2 H96711  
33 183 14.8 1460 1 EDBEIF  
34 182.5 14.8 214 2 T10737  
35 182.5 14.8 667 2 T17221  
36 182.5 14.8 1048 2 T31425  
37 180.5 14.6 280 2 T11671  
38 180 14.6 322 2 S25299  
39 180 14.6 393 2 PQ0479  
40 180 14.6 439 2 S51939  
41 180 14.6 1201 2 G86441  
42 179.5 14.5 163 2 A29356  
43 179.5 14.5 309 2 S08343  
44 179.5 14.5 369 2 S20500  
45 179.5 14.5 440 2 JC7807

ALIGNMENTS

RESULT 1

WMBEW6 capsid protein - human herpesvirus 1 (strain 17)

C:Species: human herpesvirus 1

C:Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 09-Jul-2004

C:Accession: H30084

R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Per

J. Gen. Virol. 69, 1531-1574, 1988

A:Title: The complete DNA sequence of the long unique region in the genome of herpes sim

A:Reference number: A30083; MUID:88274327; PMID:2839594

A:Accession: H30084

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-635 <MCG>

A:Cross-references: UNIPROT:P10210; GB:X14112; NID:9144536; PIDN:CAA32318.1; PID:959526

C:Genetics:

A:Gene: UL26

C:Superfamily: varicella-zoster virus gene 33 protein

C:Keywords: capsid protein

Query Match 54.2%; Score 670.5; DB 1; Length 635;  
Best Local Similarity 63.4%; Pred. No. 2.5e-33;  
Matches 144; Conservative 12; Mismatches 64; Indels 7; Gaps 4;

QY 1 AIAADROAGGLPAAAGDHGIRGSAKRREHEVEQPEYDCGRDEPRDRFFYPYGEARPEPRP 60

Db 403 AIAADROAGGQP-AAGDPSVGRSGKRRYEACPSYCDQDEPDADYPYPGEARGAPRG 461

QY 61 VDSRAARQASGPHETITATLVGAVTSLOQLAHMARHAPYGPYPVGYHHPHADTE- 119

Db 462 VDSRAARHSPTNETITATLVGAVTSLOQLAHMARHAPYGPYPVGYHHPHADTE- 521

QY 120 TPAQPPRYPAKAVYLPPIAPPGPLSGA--VPPSPYPVAVTPGAPPLHQPSPAHAH 177

Db 522 TTTTHPALCPPEAVRPPPHSPAPYGPYPVGYHHPHADTE-PCPSTOT 578

QY 178 PPPPPPGTTPPAASLPQEPAGAGALVNASSAAHVNVDTPARAAD 224

Db 579 RAPLTPFAFPATGSPQASNAEAGALVNASSAAHVNVDTPARAAD 625

RESULT 2

S52216 viral proteinase - rabies virus

C:Species: rabies virus

C:Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 21-Aug-1998

C:Accession: S52216

R:Camacho, A.; Tabaro, E.

submitted to the EMBL Data Library, June 1994

A:Reference number: S52215

A:Accession: S52216

A:Status: preliminary

A;Molecule type: genomic RNA  
A;Residues: 1-522 <CAM>  
A;Cross-references: EMBL:X79983  
A;Note: the source is designated as pseudorabies virus  
C;Superfamily: varicella-zoster virus gene 33 protein

```

Query Match      17.9%; Score 221.5; DB 2; Length 522;
Best Local Similarity 32.5%; Pred. No. 1.5e-06;
Matches 89; Conservative 22; Mismatches 78; Indels 85; Gaps 19;

Qy 1 AIAADROAGGLPAAGDGHGINGS-----AKRRHEVEQPEYDCGRDEPDRFPFYYPG 52
Db | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
274 APSLESQGLGAIVSAAMDRRRHSPPSPRPAPKRRY-----DDYAQDNAYYPG 322

Qy 53 EARPEPRPVDSSRAARQASGPHETITAIIVGAVTSLQQLAHMRA---RTHAP-YGYPYP- 107
Db | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
323 EAPP-----PRATSRV-----VSSLOREISSLRAQHVRYPPTYVAPAPQ 363

Qy 108 -----VG-PYHHPHADTETPAQPRRY-PAKAVYLPFP-----HIAP--EGPPLSGAVPP 152
Db | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
364 LLPGAVVGHPHPHF-HAAGALYPMVAQPGIHLAPPPSVAHAVVAILGLPASRCGP 422

Qy 153 -----PSYP-----PVATTEGPAPPL-----HQP-SPAHAHPP-PPPPGPTPPPA 190
Db | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
423 VAHVPAQWVFQPPVVQAQPFVAVPAAAPPPLRLQQRHAPAPVQAAPRAPASAPQFPVQ 482

Qy 191 ASLPOPEAPGAEGALVNASSAAHVVDTARAAD 224

Db 483 ASVSAPR-PTESPAPIDASSAA---VACORGAD 512

```

RESULT 3  
T05225  
extensin homolog F1715.160 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C/Accession: T05225  
R/Bevan, M.; Vitale, D.; Liguori, R.; Argirou, A.; De Simone, V.; Hoheisel, J.; Mewes,  
submitted to the Protein Sequence Database, July 1998  
A/Reference number: Z15404  
A/Accession: T05225  
A/Molecule type: DNA  
A/Residues: 1-699 <BEV>  
A/Cross-references: UNIPROT:O81765; EMBL:AL031032  
A/Experimental source: cultivar Columbia; BAC clone F1715  
C/Genetics:  
A/Map position: 4  
A/Note: F1715.160

	Query Match	17.4%; Score 215; DB 2; Length 699;	
	Best Local Similarity	32.5%; Pred. No. 4.7e-06;	
	Matches	53; Conservative 17; Mismatches 65; Indels 28; Gaps 6;	
QY	55 RPVDSRRARQASGPHEITIALVCAGVTSLQQELAHMFAHTAP-----YG 103	: ::   :	:
Dd	482 KPSEVSERPVKQPFPKESQPDFDYP-----QSPVTKRSPPPAVNSSPPPVPYS 532	::   :	:
QY	104 PYPVPVGPHHFIADTETPAOPPRY---PAKAVLPLPPIIAPPGLPSLGAVPPPSY--p 156	::   :	:
Dd	533 PPPEPVHSFPPPVHSFPPPVYSPPPPPPPVHSFPPPVFSPPPVY-SPPPPVHSGPPP 591	::   :	:
QY	157 PVAVTGCPAPPLHQGPSAHHAPPPPGFTPTTPAAALPQPEAP 199	::   :	:
Dd	592 PVHSPPPPAVSHGPPPPVHGPPPPPVYSPPPVFSPPPSOSP 634	::   :	:

RESULT 4  
WZBEC8  
68.6K capsid protein - equine herpesvirus 1 (strain Ab4p)  
C:Species: equine herpesvirus 1  
A:Note: host Equus caballus (domestic horse)  
C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text-  
C:Accession: I36798

R;Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.  
submitted to GenBank, March 1992  
A;Description: The DNA sequence of equine herpesvirus-1.  
A;Reference number: A36805  
A;Accession: I36798  
A;Molecule type: DNA  
A;Residues: 1-646 <TEL>  
A;Cross-references: UNIPROT:P28936; GB:M86664; NID:G330791;  
R;Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.  
Virology 189, 304-316, 1992  
A;Title: The DNA sequence of equine herpesvirus-1.  
A;Reference number: A41931; WJID:9229566; PMID:1318606  
A;Contents: annotation; possible protein-coding frames  
A;Note: neither amino acid nor nucleotide sequence is given  
C;Genetics:  
A;Gene: 35  
C;Superfamily: varicella-zoster virus gene 33 protein  
C;Keywords: capsid protein

```

Query Match      17.1%; Score 211; DB 1; Length 646;
Best Local Similarity 30.9%; Pred. No. 7.6e-06;
Matches 77; Conservative 25; Mismatches 55; Indels 92; Gaps 12;

Qy   1 AIAADQAAGLPAACDGHGIRGS-----AKRRHEVDEPYDCGRDEPRD 46
    ||||| : ||||| : ||||| : | : | :
Db   413 ALAADRKA---TKGSDPHVIQSQCPLSPQQERYRYARKRHWDATTRD-----DLK 463

Qy   47 FPYPYGEAREPEPVSRAAQAQSPHETITATLVCAVTSLQOELAHMEA-----RTH 99
    ||||| : ||||| : ||||| : ||||| : ||||| :
Db   464 GIYPGERSRP---GERRAGR-----PSTTIADLMGAVSSQLQVESQALRAIOTVTAAPOA 516

Qy   100 APYGPPYPVGPYHHPHADTETAQPPRYPAKAYVLPPP-----TAPPGPLSGAVPPPS 154
    ||||| : ||||| : ||||| : ||||| : ||||| :
Db   517 APAGLYKPI-----PAVPQY-SQYIQIQOHAVSAIVAPQLPGI-----PS 557

Qy   155 YPVAVTPGPAPPLHQFSPAHAHPPPPPPTPPPAASLPQEAPGAEGALVNASSAAH 214
    ||||| : ||||| : ||||| : ||||| :
Db   558 QPTQAVL-----APQVPAGEAFGS---AKVVAASTAP 586

Qy   215 VNVDTARAA 223
    ||||| : ||||| :
Db   587 QQAEQARAA 595
```

RESULT 5  
C84534  
hypothetical protein At2g15880 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: C84534  
R.;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii,  
M.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tal-  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ve-  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: C84534  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-727 <STO>  
A;Cross-references: UNIPROT:Q9XIL9; GB:AE002093; NID:G5306245; PIDN:AAD41978.1; G  
C;Genetics:  
A;Gene: At2g15880  
A;Map position: 2

```

Query Match      16.7%; Score 206; DB 2; Length 727;
Best Local Similarity 29.0%; Pred. No. 1.7e-05;
Matches 72; Conservative 16; Mismatches 86; Indels 74; Gaps 14;

Qy  2  IAAQROAGGLPAAAGDHGIRGSAKRHRHVEQPEYDCGRD-----PD 44
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db  349  IAAQDTRNCLP-----DRPKORSACECAVVTISR-VDCSKDKCAGGSOATPKSPSPVPT 403

```

```
Qy 45 RDP--PYPGEARPEPR-----PVDERRAARQASGHETTALVGAVTSLQOELAHMRA 96
Db 404 RPVHKPQPKES--QPNDPVNQSVKFRSSPPPPQOQH-----HHVHSPP 448
Qy 97 RTHAPYGYPPV---GPYHHPHADTETPAQPPY-----PAKAVYLLPP--- 137
Db 449 PASSP-PTSPVPHVTPGPKQPKESQPNDDYDQSVKFRSSPPPPVHSSPPSPSP 507
Qy 138 HIAPGPPPLSCAVPPPSY---PPVAVTPGAPPLHQ--PSPAHAHP-----PPPGGPTTP 188
Db 508 HSPPPPPVYSPPPPPVYSPPPPPVYSPPPPPVHSSPPPPVHSSPPPPVHSSPP 567
Qy 189 PAASLPQP 196
Db 568 PPVHSPP 575

RESULT 6
S06733
hydroxyproline-rich glycoprotein precursor - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C>Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 09-Jul-2004
C:Accession: S06733
R:Keller, B.; Lamb, C.J.
Genes Dev. 3, 1639-1646, 1989
A:Title: Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene
A:Reference number: S06733; MUID:90128263; PMID:2612909
A:Accession: S06733
A:Molecule type: DNA
A:Residues: 1-620 <HEL>
A:Cross-references: UNIPROT:P13983; EMBL:X13885; NID:g19866; PIDN:CAA32090.1; PID:g19867
C:Superfamily: hydroxyproline-rich glycoprotein
C:Keywords: Glycoprotein

Query Match 16.5%; Score 204; DB 2; Length 620;
Best Local Similarity 27.7%; Pred. No. 1.9e-05;
Matches 65; Conservative 21; Mismatches 81; Indels 68; Gaps 10;

Qy 28 RHEVQPEYD-----CGRDEPRDF-PVYPGEARPEPRPVDERRAARQASGHETTAL 80
Db 257 RQPOQPTYSPPPPAYASQPSPTYSPPPTYSPPPSPIYSPPPPAYSPPTPTP- 315
Qy 81 VGAVTSLQOELAHMRATHAP-----YGPYPPV-----GPYHHPHADT----- 118
Db 316 ----TFSPPPPAYSPPTYSPPPTVPLPSSPIYSPPPVYSPPPPSYSPPTYLPP 371
Qy 119 -----ETPAQPPY-----PAKAVYLLPP-----HIAPGPPPL 146
Db 372 PPPSPPPPSFSPPTVYEQSPPPPPAYSPPLFAPPTYSPPPTYSPPPTTYPQPPPLP 431
Qy 147 SGAVPPPSY-PPVAVTPGAPPLHQ--PSPAHAHPPPPPGPTPPPAASLPQPEAP 199
Db 432 TYSPPPPAYSPPTYSPPPTYSPPPPAYAQPPPPPTYSPPPPPAYSPPPSP 486

RESULT 7
S22697
extensin - Volvox carteri (fragment)
C:Species: Volvox carteri
C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Jul-2004
C:Accession: S22697; S21006
R:Ertl, H.; Hallmann, A.; Wenzl, S.; Sumper, M.
EMBO J. 11, 2055-2062, 1992
A:Title: A novel extensin that may organize extracellular matrix biogenesis in Volvox ca
A:Reference number: S22697; MUID:92289669; PMID:1600938
A:Accession: S22697
A:Molecule type: mRNA
A:Residues: 1-464 <HAL>
A:Cross-references: UNIPROT:Q41645; EMBL:X65165; NID:g21991; PIDN:CAA46283.1; PID:g21992
C:Keywords: glycoprotein

Query Match 16.2%; Score 200; DB 2; Length 464;
Best Local Similarity 33.1%; Pred. No. 2.6e-05;
```

```
Matches 57; Conservative 10; Mismatches 71; Indels 34; Gaps 7;

Qy 48 PYTGEARPEPRPVDERRAARQASGHETTALVGAVTSLQOELAHMRATHAPYGYPP 107
Db 232 PPPARVSSSPPATRPPRRITSPSVLTA-----SPPLPKTSP-PPPR 277
Qy 108 VGPVHHPHADTETPAQPPRY-----PAKAVYLLPPPHIAP-----PGPPLSG-----AVP 151
Db 278 VPSPPPPPVAVSPPPPPPPVSPSPPPPPVSPPPPPPPPPPPPPPPPPPPPPPPSP 337
Qy 152 PPSYPPVAVTPGAPPLHQPSPAH--AHPPPPPP--GTPPPPAASLPQPEAP 199
Db 338 PPSPPPPRPSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 389

RESULT 8
T06291
extensin homolog T9E8.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
C:Accession: T06291
R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schuel
submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15588
A:Accession: T06291
A:Molecule type: DNA
A:Residues: 1-760 <BEV>
A:Cross-references: UNIPROT:Q9TK05; EMBL:AL049608
A:Experimental source: cultivar Columbia; BAC clone T9E8
C:Genetics:
A:Map position: 4
A:Note: T9E8.80

Query Match 16.0%; Score 198; DB 2; Length 760;
Best Local Similarity 42.6%; Pred. No. 5.2e-05;
Matches 46; Conservative 3; Mismatches 47; Indels 12; Gaps 3;

Qy 102 YGPYPPVGPYHHPHADTETPAQPPRYPAKAVYLLPPPHIAPGPP--LSGAVPPPSYPPVA 159
Db 449 YSPPPPPPPPPPPVYSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 508
Qy 160 VTPGAPPLHQ-----PSPAH-----AHPPPPPPGPTPPPAASLPQPE 197
Db 509 VYSPPPPPVYSSPPSPSPAPTPVCTRPDPHSPPPPPQFSPPPE 556

RESULT 9
C96623
hypothetical protein F23H11.22 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: C96623
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C96623
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-929 <STO>
A:Cross-references: UNIPROT:Q9XIE0; GB:AE005173; NID:g5080823; PIDN:AAD39332.1; GSPDB:G
C:Genetics:
A:Map position: 1

Query Match 15.8%; Score 196; DB 2; Length 929;
```







This page blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2005, 12:06:17 ; Search time 44 Seconds  
(without alignments)  
380.032 Million cell updates/sec

Title: US-10-623-429-9\_COPY\_404\_627  
Perfect score: 1237  
Sequence: 1 AIAADRQAGGLPAAAGDHGI.....ALVNASSAAHVNDTARAAD 224

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues  
Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pap.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pap.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pap.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pap.\*  
5: /cgn2\_6/ptodata/1/iaa/PTCUS\_COMB.pap.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	670.5	54.2	635	4	US-08-176-320-2
2	660.5	53.4	635	1	Sequence 2, Appli
3	190.5	15.4	553	4	US-07-832-855-2
4	190	15.4	324	4	US-09-949-016-7961
5	189.5	15.3	498	4	Sequence 7664, Ap
6	184.5	14.9	559	4	Sequence 7108, Ap
7	184.5	14.9	567	4	Sequence 2, Appli
8	184	14.9	297	2	Sequence 10952, A
9	184	14.9	297	3	Sequence 6, Appli
10	179	14.5	511	4	Sequence 2904, Ap
11	178.5	14.4	142	4	Sequence 24873, A
12	178	14.4	104	4	Sequence 235, App
13	178	14.4	707	4	Sequence 278, App
14	178	14.4	707	4	Sequence 993, App
15	178	14.4	735	4	Sequence 10120, A
16	177.5	14.3	581	4	Sequence 978, Ap
17	177	14.3	214	1	Sequence 4, Appli
18	177	14.3	271	4	Sequence 21050, A
19	177	14.3	971	4	Sequence 19531, A
20	176.5	14.3	1274	3	Sequence 2, Appli
21	175.5	14.2	506	4	Sequence 11282, A
22	175.5	14.2	684	4	Sequence 9, Appli
23	175	14.1	288	4	Sequence 42632, A
24	175	14.1	503	4	Sequence 2, Appli
25	175	14.1	507	4	Sequence 24, Appli
26	173.5	14.0	1248	2	Sequence 2, Appli
27	173.5	14.0	1248	3	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-176-320-2  
; Sequence 2, Application US/08176320  
; Patent No. 6410704  
; GENERAL INFORMATION:  
; APPLICANT: Roizman, Bernard  
; APPLICANT: Liu, Fenyong  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
; PREPARATION AND USE OF A HERPES PROTEASE  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Alice O. Martin  
; STREET: 321 No. 6410704th Clark Street, Suite 800  
; CITY: Chicago  
; STATE: ILL  
; COUNTRY: USA  
; ZIP: 60610  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/176,320  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/705,814  
; FILING DATE: 24-MAY-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cooley, Ronald B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312)744-0090  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 635 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-176-320-2

Query Match 54.2%; Score 670.5; DB 4; Length 635;  
Best Local Similarity 63.4%; Pred. No. 5.1e-44;  
Matches 144; Conservative 12; Mismatches 64; Indels 7; Gaps 4;

QY 1 AIAADRQAGGLPAAAGDHGIRGSAKRHRHEVQPEYDCGRDEPDRDFPYPGGEARPEPRP 60  
DB 403 AIAADRQAGGLPAAAGDHGIRGSAKRHRHEVQPEYDCGRDEPDRDFPYPGGEARPEPRP 461  
QY 61 VDSRRRAARQASGPHETITATVAVGAVTSLOQELAHMRARTHAPYGPVPFVGPHHPHADTE- 119

Db	462	VDSRRAARHSPTNETITATLMGAVTSLQOELAHMRARTSAPYGMTTPVAHYRPQVGEPEP	521
Qy	120	TPAQPPRYPKAVYLPPIHPIAPPGLPSGA--VPPPSYPPVAVTPGPAPLLHQPSPAHAH	177
Db	522	TTTHPALCPPEAVYRPPSHSAPYGPQGPAASHATPPYAPAAACPPGPPPP---PCPSTQT	578
Qy	178	PPPPPPGPTPPPAASLQPOEAPGABAGALVNASSAAHVNVDTARAAD	224
Db	579	RAPLPTPEAFPPAATGSGQPEASNAEAGALVNASSAAHVVDVDTARAAD	625
RESULT 2			
US-07-832-855-2			
; Sequence 2, Application US/07832855			
; Patent No. 5478727			
; GENERAL INFORMATION:			
; APPLICANT: Roizman, Bernard			
; APPLICANT: Liu, Fenyong			
; TITLE OF INVENTION: Methods and Compositions of a			
; TITLE OF INVENTION: Preparation and Use of A Herpes Protease			
; NUMBER OF SEQUENCES: 15			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: ARNOLD, WHITE & DURKEE			
; STREET: 321 No. 5478727th Clark Street, Suite 800			
; CITY: Chicago			
; STATE: Illinois			
; COUNTRY: USA			
; ZIP: 60610			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: Patent In Release #1.0, Version #1.25			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/07/832.855			
; FILING DATE: 19920207			
; CLASSIFICATION: 435			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Cooley, Ronald B.			
; REGISTRATION NUMBER: 27,187			
; REFERENCE/DOCKET NUMBER: ARCD045			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (312) 744-0090			
; TELEFAX: (312) 245-4961			
; INFORMATION FOR SEQ ID NO: 2:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 635 amino acids			
; TYPE: AMINO ACID			
; TOPOLOGY: linear			
; MOLECULE TYPE: protein			
US-07-832-855-2			
Query Match 53.4%; Score 660.5; DB 1; Length 635;			
Best Local Similarity 62.6%; Pred. No. 3e-43;			
Matches 142; Conservative 12; Mismatches 66; Indels 7; Gaps 4			
Qy	1	AIAADRAQAGLPPAAAGDHGIRGSAKRRRHEVEQPEYDCGRDEPDRDPFYYPGEARPERPP	60
Db	403	AIAADRAQAGQP--AAGDGVGRSGKRRRYEAGPSSEYCDQEPDADYPYPGEARGAPRG	461
Qy	61	VDSRRAARQSGPHETITATLVGAVTSLQOELAHMRARTHAPYGPYPVPVHHPHADTE-	119
Db	462	VDSRRAARHSPTNETITATLMGAVTSLQOELAHMRARTSAPYGMTTPVAHYRPQVGEPEP	521
Qy	120	TPAQPPRYPKAVYLPPIHPIAPPGLPSGA--VPPPSYPPVAVTPGPAPLLHQPSPAHAH	177
Db	522	TTTHPALCPPEAVYRPPSHSAPYGPQGPAASHATPPYAPAAACPPGPPPP---PCPSTQT	578
Qy	178	PPPPPPGPTPPPAASLQPOEAPGABAGALVNASSAAHVNVDTARAAD	224
Db	579	RAPLPTPEAFPPPTTGGSQPEASNAEAGALVNASSAAHVVDVDTARAAD	625

Db 130 APSGPPGPPQLPPAPGVPHPVHP-----PASGVHPAPGVPHPAPGVPVH 183  
Qy 151 PPS--YPPVAVTGPAPPLHQPSAHP-----PPPGPTPPPAASLPQEPAGAE 203  
Db 184 PPTSGVHPAPGVPHPAPGVPHPAPG--VHPAPGVPHPAPGVPHPAPG--HPQAPGVHP 241  
Qy 204 GA 205  
Db 242 AA 243

RESULT 5  
US-09-949-016-7108  
; Sequence 7108, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7108  
; LENGTH: 498  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-7108

Query Match 15.3%; Score 189.5; DB 4; Length 498;  
Best Local Similarity 32.1%; Pred. No. 4.3e-07;  
Matches 54; Conservative 12; Mismatches 45; Indels 57; Gaps 9;  
Qy 43 PDRDFYYGEPPEPRVDSRRARQASGPHETITLVGAVTSLQQLAHMRARTHAPY 102  
Db 270 PPAEFY-----PVDNRGSLA-GPKRSW-----SPSHPP- 301  
Qy 103 GPYPVGVYHHPADTETPAQPRYPKAVYLPPIHAPPGLSGAVPPSPYV----- 158  
Db 302 -PAPPLG-----SPGP-----KPGFAPPAPPPPPPMIGIPPP--PPVGRGS 343  
Qy 159 -AVTPGAPPLHQPSAHPHPPPPGPT-----PPPAASLPQEPAP 199  
Db 344 PGTTPPPSPSPHPHPDFAATPPPPPPPAADYPTLPPPLSQPTGGAP 391

RESULT 6  
US-10-116-370-2  
; Sequence 2, Application US/10116370  
; Patent No. 6664068  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; TITLE OF INVENTION: PABLO, A POLYPEPTIDE THAT INTERACTS WITH BCL-XL, AND  
; TITLE OF INVENTION: USES RELATED THERETO  
; FILE REFERENCE: AM100012-D2  
; CURRENT APPLICATION NUMBER: US/10/116,370  
; CURRENT FILING DATE: 2002-04-04  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 559  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-116-370-2

Query Match 14.9%; Score 184.5; DB 4; Length 559;  
Best Local Similarity 31.5%; Pred. No. 1.2e-06;  
Matches 63; Conservative 13; Mismatches 79; Indels 45; Gaps 9;  
Qy 55 RP-EPRVPDSRRARQASGPHETITLVGAVTSLQQLAHMRARTHAPYGVPPVGVYH 113  
Db 274 RPHEPPPPPMHAGDAKPIPTCISSTGLIENRQSPA--TGRTPVFSVTPPPPPPL 331  
Qy 114 PHADTETPAQPRYPKAVYLPPIHAPPGP-----LSGAVPPSPYV-----PVAVTGP 164  
Db 332 PSALSTSSLR-----ASMTSTPPPPVPPPPPPATLQAPAVPPPPAPLQIAPGVLPAP 386  
Qy 165 ---APPLHQPSAHP-----PPPPPGTTPPPAASLPQEPAPG 200  
Db 387 PPIAPPLVQSPFPVARAAPVCETVPVHPLPQGEVQGLPPPPPPPLPPPGI---RPSSP- 442  
Qy 201 AEAGALVNASSAAHVNVDTA 220  
Db 443 VVTVALAHPPSGLHPTPSTA 462

RESULT 7  
US-09-949-016-10952  
; Sequence 10952, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10952  
; LENGTH: 567  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-10952

Query Match 14.9%; Score 184.5; DB 4; Length 567;  
Best Local Similarity 31.5%; Pred. No. 1.2e-06;  
Matches 63; Conservative 13; Mismatches 79; Indels 45; Gaps 9;  
Qy 55 RP-EPRVPDSRRARQASGPHETITLVGAVTSLQQLAHMRARTHAPYGVPPVGVYH 113  
Db 282 RPHEPPPPPMHAGDAKPIPTCISSTGLIENRQSPA--TGRTPVFSVTPPPPPPL 339  
Qy 114 PHADTETPAQPRYPKAVYLPPIHAPPGP-----LSGAVPPSPYV-----PVAVTGP 164  
Db 340 PSALSTSSLR-----ASMTSTPPPPVPPPPPPATLQAPAVPPPPAPLQIAPGVLPAP 394  
Qy 165 ---APPLHQPSAHP-----PPPPPGTTPPPAASLPQEPAPG 200  
Db 395 PPIAPPLVQSPFPVARAAPVCETVPVHPLPQGEVQGLPPPPPPPLPPPGI---RPSSP- 450  
Qy 201 AEAGALVNASSAAHVNVDTA 220  
Db 451 VVTVALAHPPSGLHPTPSTA 470

RESULT 8  
US-08-580-545B-6  
; Sequence 6, Application US/08580545B  
; Patent No. 5932713  
; GENERAL INFORMATION:  
; APPLICANT: Yoshihisa, Kasukabe



```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...511
; SEQUENCE DESCRIPTION: SEQ ID NO: 2904:
US-09-107-433-2904

Query Match      14.5%; Score 179; DB 4; Length 511;
Best Local Similarity 34.3%; Pred. No. 2.8e-06;
Matches 61; Conservative 9; Mismatches 64; Indels 44; Gaps 11;

Qy 47 FPYVGEARPRVDSSRAARQASGHETITLVGAVTSIQLQELAHM----RARTH--A 100
Db 18 FPHPP--YTHAPPADPIATSPFVKRGLTTHAIRA-----AMISLTPRSSHQLP 67

Qy 101 PYGYPVPV-GPYHHPHADTETPAQ--PPRYPAKAVYLPPIPHIAP-----PPPLSGAVPP 152
Db 68 PTHPTPPVPGP-----PPQPRPHPPPPV---PPHAAFLHAANPPRASNSPPP 115

Qy 153 PSYPPVAVTPGAPPLHQ-----PSPAHAPPPPPP--GPTPPPAASLPOPEAPGAE 202
Db 116 PPPPAAQQPHAPPRNPPHPGTPHRRHPTPPPTTPPTTPPAHTGPTTPTTNE 173

RESULT 11
US-09-252-991A-24873
; Sequence 24873, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24873
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24873

Query Match      14.4%; Score 178.5; DB 4; Length 142;
Best Local Similarity 46.8%; Pred. No. 8.2e-07;
Matches 37; Conservative 2; Mismatches 33; Indels 7; Gaps 2;

Qy 121 PAQPPRYPAKAVYLPPIPHIAPGPPPLSGAVPPSPYPPVAVTPGAPPLHQPSPAHAHPPP 180
Db 60 PVPPVPVPPSPSPPPPPVPPSPVPG-VPPSPSP-----PVPPSPSPSPPPPPVPPS 112

Qy 181 PPPGTPPPPAASLPOPEAP 199
Db 113 PPPSPPPVPPSPPPGPPPP 131

RESULT 12
US-09-547-693-235
; Sequence 235, Application US/09547693
; Patent No. 6639050
; GENERAL INFORMATION:
; APPLICANT: Kieliszewski, Marcia
; TITLE OF INVENTION: Synthetic Genes for Plant Gums and Other Hydroxyproline-Rich
; FILE REFERENCE: Glycoproteins
; CURRENT APPLICATION NUMBER: OHU-04089

```

```

; CURRENT APPLICATION NUMBER: US/09/547,693
; CURRENT FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 236
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 235
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Artificial/Unknown
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Synthetic
US-09-547-693-235

Query Match      14.4%; Score 178; DB 4; Length 104;
Best Local Similarity 40.0%; Pred. No. 6.5e-07;
Matches 42; Conservative 2; Mismatches 51; Indels 10; Gaps 3;

Qy 95 RARTHAPYGVPPVGPYHHPHADTETPAQPPRYPAKAVYLPPIPHIAPGPPPLSGAVPPS 154
Db 5 RASPPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 56

Qy 155 YPPVAVTPGAPPLHQPSPAHAHPPPPPPGTPPPPAASLPOPEAP 199
Db 57 SPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 99

RESULT 13
US-09-919-039-278
; Sequence 278, Application US/09919039
; Patent No. 6727066
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 278
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6727066 2457215CD1
US-09-919-039-278

Query Match      14.4%; Score 178; DB 4; Length 707;
Best Local Similarity 39.7%; Pred. No. 4.8e-06;
Matches 46; Conservative 6; Mismatches 40; Indels 24; Gaps 6;

Qy 103 GPYPVPVGPYHHPHADTETPAQPPRYPAKAVYLPPIPHIAPGPP-PLSGAVPPP-----SYP 156
Db 55 GKPKPIPP-PPPHQQQQPP--PPQQPPP--QQPPHQPPIPHIAPHQQQQPPPPQDSSKP 109

Qy 157 PVAVTPGAPPLHQPSPAHAHPPP-----PPPGTTPPPPAASLPOPEAP 199
Db 110 VVAQPGPAPGVGSAPPASSAPPATPTTSGAPGSGGPTTTPPPAVTSAPPAGAP 165

RESULT 14
US-09-538-092-993
; Sequence 993, Application US/09538092
; Patent No. 6753114
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29

```

```
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 993
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P23246
US-09-538-092-993

Query Match      14.4%; Score 178; DB 4; Length 707;
Best Local Similarity 39.7%; Pred. No. 4.8e-06;
Matches 46; Conservative 6; Mismatches 40; Indels 24; Gaps 6;

QY 103 GPYPVGVYHHADTETPAQPRYPYAKAVYLPPIHAPPGP-PLSGAVPPP-----SYP 156
Db 55 GKPPPIPP-PPPHQQQQP--PPQQPPP--QQPPPHQPPPHQPPPPPPQDSSKP 109
QY 157 PVAVTGPAPPLHQPSFAHAHPPP-----PPPGTTPPPAASLPQPEAP 199
Db 110 VVAQGGPAPGVGSAPPASSAPPATPTTSGAPPGSGGPTTTPPPAVTSAPPGAP 165

RESULT 15
US-09-949-016-10120
; Sequence 10120, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10120
; LENGTH: 735
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10120

Query Match      14.4%; Score 178; DB 4; Length 735;
Best Local Similarity 39.7%; Pred. No. 5e-06;
Matches 46; Conservative 6; Mismatches 40; Indels 24; Gaps 6;

QY 103 GPYPVGVYHHADTETPAQPRYPYAKAVYLPPIHAPPGP-PLSGAVPPP-----SYP 156
Db 83 GKPPPIPP-PPPHQQQQP--PPQQPPP--QQPPPHQPPPHQPPPPPPQDSSKP 137
QY 157 PVAVTGPAPPLHQPSFAHAHPPP-----PPPGTTPPPAASLPQPEAP 199
Db 138 VVAQGGPAPGVGSAPPASSAPPATPTTSGAPPGSGGPTTTPPPAVTSAPPGAP 193
```

Search completed: July 27, 2005, 12:23:14  
Job time : 46 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2005, 12:21:48 ; Search time 155 Seconds  
(without alignments)  
562.157 Million cell updates/sec

Title: US-10-623-429-9\_COPY\_404\_627  
Perfect score: 1237  
Sequence: 1 AIAADRQAGGLPAAAGDHGI.....ALVNASSAAHVNDTARAAD 224

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1741741 seqs, 388992284 residues 1741741  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US11A\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW PUB.pep.\*  
21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW PUB.pep.\*  
22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

*Result No.	Score	Query Match	Length	ID	Description
1	1237	100.0	637	15	US-10-623-429-9
2	670.5	54.2	635	14	US-10-214-932-86
3	222	17.9	275	16	US-10-425-115-315706
4	221.5	17.9	376	16	US-10-767-701-42767
5	219.5	17.7	342	16	US-10-437-963-112806
6	218.5	17.7	241	15	US-10-424-599-224421
7	217	17.5	206	15	US-10-424-599-224421
8	215.5	17.4	187	16	US-10-425-115-352789
9	214	17.3	149	16	US-10-425-115-252326
10	213.5	17.3	204	15	US-10-424-599-221495
11	213.5	17.3	298	16	US-10-437-963-147019
12	1237	100.0	637	15	US-10-623-429-9
13	670.5	54.2	635	14	US-10-214-932-86
14	222	17.9	275	16	US-10-425-115-315706
15	221.5	17.9	376	16	US-10-767-701-42767
16	219.5	17.7	342	16	US-10-437-963-112806
17	218.5	17.7	241	15	US-10-424-599-224421
18	217	17.5	206	15	US-10-424-599-224421
19	215.5	17.4	187	16	US-10-425-115-352789
20	214	17.3	149	16	US-10-425-115-252326
21	213.5	17.3	204	15	US-10-424-599-221495
22	213.5	17.3	298	16	US-10-437-963-147019

12	213.5	17.3	639	15	US-10-369-493-3962	Sequence 3962, Ap
13	213	17.2	147	15	US-10-424-599-154677	Sequence 154677,
14	213	17.2	178	15	US-10-424-599-264218	Sequence 264218,
15	213	17.2	223	16	US-10-425-115-230534	Sequence 230534,
16	212	17.1	140	15	US-10-424-599-276630	Sequence 276630,
17	211.5	17.1	296	15	US-10-424-599-224246	Sequence 224246,
18	211	17.1	107	15	US-10-424-599-234553	Sequence 234553,
19	211	17.1	329	17	US-10-626-832-41	Sequence 41, Appl
20	211	17.1	497	16	US-10-437-963-122257	Sequence 122257,
21	211	17.1	646	17	US-10-626-832-42	Sequence 42, Appl
22	210	17.0	149	15	US-10-424-599-278484	Sequence 278484,
23	210	17.0	273	16	US-10-425-115-330708	Sequence 330708,
24	210	17.0	409	16	US-10-437-963-140753	Sequence 140753,
25	209	16.9	183	16	US-10-425-115-239672	Sequence 239672,
26	209	16.9	327	16	US-10-425-115-276286	Sequence 276286,
27	208.5	16.9	187	16	US-10-425-115-300973	Sequence 300973,
28	208.5	16.9	204	16	US-10-437-963-119585	Sequence 119585,
29	208.5	16.9	235	16	US-10-437-963-171176	Sequence 171176,
30	208	16.8	165	16	US-10-767-701-35865	Sequence 35865, A
31	208	16.8	280	16	US-10-425-115-233956	Sequence 233956,
32	207.5	16.8	236	16	US-10-425-115-318599	Sequence 318599,
33	207	16.7	263	16	US-10-437-963-116465	Sequence 116465,
34	206.5	16.7	180	16	US-10-425-115-229737	Sequence 229737,
35	206.5	16.7	431	16	US-10-437-963-204963	Sequence 204963,
36	206	16.7	189	15	US-10-424-599-157515	Sequence 157515,
37	206	16.7	229	15	US-10-424-599-207316	Sequence 207316,
38	205	16.6	252	16	US-10-437-963-151853	Sequence 151853,
39	205	16.6	299	16	US-10-437-963-134054	Sequence 134054,
40	204.5	16.5	133	16	US-10-425-115-229322	Sequence 229322,
41	204.5	16.5	400	16	US-10-425-115-278048	Sequence 278048,
42	204	16.5	125	16	US-10-425-115-278144	Sequence 278144,
43	204	16.5	178	16	US-10-425-115-321259	Sequence 321259,
44	204	16.5	208	16	US-10-437-963-103915	Sequence 103915,
45	204	16.5	267	16	US-10-425-115-238339	Sequence 238339,

ALIGNMENTS

RESULT 1  
US-10-623-429-9  
; Sequence 9, Application US/10623429  
; Publication No. US20040072152A1  
; GENERAL INFORMATION:  
; APPLICANT: ZHI LIU  
; APPLICANT: LAWRENCE COREY  
; TITLE OF INVENTION: RAPID, EFFICIENT PURIFICATION OF  
; TITLE OF INVENTION: HSV-SPECIFIC T-LYMPHOCYTES AND HSV ANTIGENS IDENTIFIED VIA  
; FILE REFERENCE: 30967.11USU1  
; CURRENT APPLICATION NUMBER: US/10/623,429  
; CURRENT FILING DATE: 2003-07-18  
; PRIOR APPLICATION NUMBER: 60/396,791  
; PRIOR FILING DATE: 2002-07-18  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 637  
; TYPE: PRT  
; ORGANISM: Herpes Simplex Virus 2  
US-10-623-429-9

Query Match	100.0%	Score 1237;	DB 15;	Length 637;
Best Local Similarity	100.0%	Pred. No. 1.3e-72;		
Matches 224;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AIAADRQAGGLPAAAGDHGIRGSAKRHEVEQPYDCGRDEPDRDFPYYPGEARPEPRP	60	
Db	404	AIAADRQAGGLPAAAGDHGIRGSAKRHEVEQPYDCGRDEPDRDFPYYPGEARPEPRP	463	
QY	61	VDSRRARQAGPHTTITVALGVATVLSLOELAHMARHAPYGPYPVPVGHPHADTET	120	

```
Db 464 VDSRRARQASGPHETITLVGAVTSLQQLAHMRARTHAPYGYPPVGYHHPHADTET 523
Qy 121 PAQPPRYPAKAVLPPPHIAPGGLSCGAVPPSPVAVTPGAPPLHQPSPAHAPPP 180
Db 524 PAQPPRYPAKAVLPPPHIAPGGLSCGAVPPSPVAVTPGAPPLHQPSPAHAPPP 583
Qy 181 PPGPTPPPAASLPQEPAPGAEGALVNASSAAHVNDTARAAD 224
Db 584 PPGPTPPPAASLPQEPAPGAEGALVNASSAAHVNDTARAAD 627

RESULT 2
US-10-214-932-86
; Sequence 86, Application US/10214932
; Publication No. US20030100707A1
; GENERAL INFORMATION:
; APPLICANT: HWANG, Inhwan
; APPLICANT: KIM, Dae Heon
; APPLICANT: LEE, Yong Jik
; TITLE OF INVENTION: SYSTEM FOR DETECTING PROTEASE
; FILE REFERENCE: APB02/US
; CURRENT APPLICATION NUMBER: US/10/214,932
; CURRENT FILING DATE: 2002-08-08
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 86
; LENGTH: 635
; TYPE: PRT
; ORGANISM: Human herpesvirus 1
US-10-214-932-86

Query Match 54.2%; Score 670.5; DB 14; Length 635;
Best Local Similarity 63.4%; Pred. No. 9.3e-36;
Matches 144; Conservative 12; Mismatches 64; Indels 7; Gaps 4;

Qy 1 AIAADQAGGLPAAAGDHGIRGSAKRRRHEVEQPEYDCGRDEPDROPFPYVGEARPEPRP 60
Db 403 AIAADQAGGQP--AAGDPGVGSGKRRRYEAGFSVCDDQEDADYYPYVGEARGPRG 461
Qy 61 VDSRRARQASGPHETITLVGAVTSLQQLAHMRARTHAPYGYPPVGYHHPHADTET- 119
Db 462 VDSRRARHSFGTNETITLTMGAVTSLQQLAHMRARTSAPYGMVTPVAHYRQVGEPEP 521
Qy 120 TPAQPPRYPAKAVLPPPHIAPGGLSCG--VPPPSYPPVAVTPGAPPLHQPSPAHAH 177
Db 522 TTHPALCPPEAVYRPPPHSAPYGPQGPASHAPTTPYAPACPPGPPPP---PCPSTQT 578
Qy 178 PPPPPPGPTPPPAASLPQEPAPGAEGALVNASSAAHVNDTARAAD 224
Db 579 RAPLTPAPPPAATGSGQPEASNAEAGALVNASSAAHVNDTARAAD 625

RESULT 3
US-10-425-115-315706
; Sequence 315706, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 315706
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure

Db 464 VDSRRARQASGPHETITLVGAVTSLQQLAHMRARTHAPYGYPPVGYHHPHADTET 523
Qy 121 PAQPPRYPAKAVLPPPHIAPGGLSCGAVPPSPVAVTPGAPPLHQPSPAHAPPP 180
Db 524 PAQPPRYPAKAVLPPPHIAPGGLSCGAVPPSPVAVTPGAPPLHQPSPAHAPPP 583
Qy 181 PPGPTPPPAASLPQEPAPGAEGALVNASSAAHVNDTARAAD 224
Db 584 PPGPTPPPAASLPQEPAPGAEGALVNASSAAHVNDTARAAD 627

Query Match 17.9%; Score 221.5; DB 16; Length 275;
Best Local Similarity 28.6%; Pred. No. 6.3e-07;
Matches 67; Conservative 10; Mismatches 83; Indels 74; Gaps 7;

Qy 1 AIAADQAGGLPAAAGDHGI-----RGSAKRRRHEVEQPEYDCGR-----D 41
Db 1 SLCPGRRAGG---GGRHGFPRRALPGDCGAQRRLRYIRRRRRFRGGAAGGLPRSQSPGD 57
Qy 42 EPDRD-----FPYVGEARPEPRVDSRRARQASGPHETITLVGAVTSLQOE 90
Db 58 EPPREKVPKTYPSXTPHARSDTPPXPPD----- 87
Qy 91 LAHMRARTHAPYGPVPPVGYHHPHADTETPAQPPRYPAKAVLPPPHIAPGGLSCG 149
Db 88 -----HAPPPPLPPGAPPEPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 139
Qy 150 VPPPSYPPVAVTPGAPPLHQPSPAHAPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 203
Db 140 RPPPP--PPGPPPPPPAPXPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPRA 191

RESULT 4
US-10-767-701-42767
; Sequence 42767, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 42767
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C17099_1.pep
US-10-767-701-42767

Query Match 17.9%; Score 221.5; DB 16; Length 376;
Best Local Similarity 34.1%; Pred. No. 9.1e-07;
Matches 71; Conservative 4; Mismatches 54; Indels 79; Gaps 11;

Qy 4 ADRQAG-GLPAAAGDHGIRGSAKRRRHEVEQPEYDCGRDEPDROPFPYVGEARPEPRVD 62
Db 16 APRPFGGLPASA---CRGPRRGGRGPRGP-----APPPPPPG 52
Qy 63 SRRAA--RQASGPHETITLVGAVTSLQQLAHMRARTHAPYGYPP-----VGYHHPHA 116
Db 53 GPPAAGPRPAPGP-----PPCGPPPPPPAPRAAGFGGPP-- 84
Qy 117 DTETPAQPPRYPAKAVLPPPHIAPGGLSCGAVPPSPVAVTPGAPPLHQPSPAH 176
Db 85 -----PAPPPPPPP-----PPPPAPPPPPPP-----PPPP-----PPPPPPPPPPPPPPPPAD--- 124
Qy 177 HPPPPPPPGTTPPAASLPQEPAPGAEGAG 204
Db 125 -PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPGGRG 151

RESULT 5
US-10-437-963-112806
; Sequence 112806, Application US/10437963
```

```

RESULT 6
US-10-424-599-163337
; Sequence 163337, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 163337
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:

```

RESULT 8  
US-10-425-115-352789  
; Sequence 352789, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovacic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei

```
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 352789
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(187)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_84911C.1.pep
US-10-425-115-352789

Query Match          17.4%; Score 215.5; DB 16; Length 187;
Best Local Similarity 36.8%; Pred. No. 1.2e-06;
Matches 57; Conservative 3; Mismatches 66; Indels 29; Gaps 6;

QY 48 PYPFGARPRFRVDSRRARQASGPHETITALTALVGAVTSLQBELAHMRARTHAPYG---P 104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 PLTPAPAPPPPPAP---PLTPATAPHT-----RPLPHSR-----PIGTTPP 50

QY 105 YPPVGPVHHPHADTETAQPRPRPAKAVYLPDPHIAPPGPPSLGAVPPPSYPVAVTPGP 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 PPPTPRLRPPAPPPXPTTPTTLPPLFSPPPPPPPPPPPPPPPPS---PPPPPPPPPPPPHP 107

QY 165 APPLHQPSPAHAPPPPPPPPPGPTTTPPAASLPQPEAP 199
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 PPELPPPPP-----PPPPPPPPPPPPPPRXPBPPPPPP 138

RESULT 9
US-10-425-115-252326
; Sequence 252326, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 252326
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(149)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_161704C.1.pep
US-10-425-115-252326

Query Match          17.3%; Score 214; DB 16; Length 149;
Best Local Similarity 36.4%; Pred. No. 1.2e-06;
Matches 52; Conservative 4; Mismatches 63; Indels 24; Gaps 5;

QY 58 PRVDSRRARQASGPHETITALTALVGAVTSLQBELAHMRARTHAPYGPY--PPVGPVHHH 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25 PPPPD-----HQNKPHNHPT-----PHSKTTQTTPPPPHPPPPSTTHHR 66

QY 116 ADTETAQPRYPYPAKAVYLPDPHIAPPGP--PLSGAVPPPSYPVAVTPGAPPLHQSP 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 NPPIPTPPXPBPPPPXPPPPHAPPPPTTTPPNHTPPPPXPPPHHTPHPPPTTTPPP 126
```

```
QY 174 AHAHPPP--PPGPTTTPPAASLP 194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 PHPHQTHTPTTPTTTPPPPPPPP 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-10-424-599-221495
; Sequence 221495, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 221495
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(204)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_42039C.1.pep
US-10-424-599-221495

Query Match          17.3%; Score 213.5; DB 15; Length 204;
Best Local Similarity 44.6%; Pred. No. 1.7e-06;
Matches 50; Conservative 0; Mismatches 49; Indels 13; Gaps 4;

QY 101 PYGPPPVGPPYHHPH-----ADTETAQPRYPYPAKAVYLPDPHIA--PPGPPPLSGA 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13 PSGFAPPPGPPPPXPPSPGPPPPGXPLPPPPGPPPLGAPLPGPPPPGAPPPGPPPPGP 72

QY 150 VPPPSYPPVAVTPGPAPPLHQPSPAHAHP--PPPGPTTTPPAASLPQPEAPG 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 73 TTPPGPPPPPLGCPAPLPGSPPPGAPPPGPPPPGPPPPGPPPPGPP--PXPPLPG 123

RESULT 11
US-10-437-963-147019
; Sequence 147019, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 147019
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(298)
; OTHER INFORMATION: unsure at all Xaa locations
```

Query Match	17.2%	Score 213;	DB 15;	Length 178;
Best Local Similarity	45.0%	Pred. No. 1.6e-06;		
Matches	45;	Conservative	1;	Mismatches 48;
			Indels	6; Gaps 2;
Seq	101	pvgpvppvgpyhphadtet-paopprrpakavylppphiiappgpplsgavppssypvva		
		159		

Search completed: July 27, 2005, 12:35:16  
Job time : 156 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 27, 2005, 12:32:44 ; Search time 41 Seconds  
(without alignments)  
16.386 Million cell updates/sec

Title: US-10-623-429-9\_COPY\_475\_483

Perfect score: 48

Sequence: 1 GPHETITAL 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/ECTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	75.0	50	3	US-09-346-860-9
2	36	75.0	50	4	US-09-735-685-9
3	36	75.0	79	3	US-08-905-223-493
4	35	72.9	58	4	US-10-127-641-5
5	35	72.9	91	4	US-09-248-796A-19847
6	35	72.9	748	2	US-08-997-080-154
7	35	72.9	748	2	US-08-997-362-154
8	35	72.9	748	3	US-09-095-855-154
9	35	72.9	748	3	US-09-324-542-154
10	35	72.9	748	4	US-09-205-426-154
11	35	72.9	754	3	US-09-005-180A-3
12	34	70.8	102	4	US-09-513-999C-6366
13	34	70.8	154	4	US-09-615-192A-311
14	34	70.8	300	4	US-09-134-000C-5791
15	34	70.8	699	4	US-09-252-991A-17077
16	34	70.8	717	4	US-09-543-681A-5706
17	34	70.8	762	4	US-09-107-532A-4275
18	34	70.8	2584	3	US-08-936-135-4
19	34	70.8	2588	3	US-08-936-135-2
20	33	68.8	81	4	US-09-107-532A-7025
21	33	68.8	364	4	US-09-252-991A-22552
22	33	68.8	401	4	US-09-248-796A-19514
23	33	68.8	635	1	US-07-832-855-2
24	33	68.8	635	4	US-08-176-320-2
25	33	68.8	726	4	US-09-489-039A-8750
26	33	68.8	1009	4	US-09-762-724-10
27	33	68.8	1017	4	US-09-762-724-12

28	33	68.8	1023	4	US-09-762-724-14	Sequence 14, Appl
29	33	68.8	1027	4	US-09-762-724-8	Sequence 8, Appl
30	32	66.7	108	4	US-09-513-999C-5317	Sequence 5317, Ap
31	32	66.7	277	3	US-09-007-484-4	Sequence 4, Appl
32	32	66.7	277	3	US-09-309-682-4	Sequence 4, Appl
33	32	66.7	305	3	US-09-335-409-22	Sequence 22, Appl
34	32	66.7	305	3	US-09-568-102-22	Sequence 22, Appl
35	32	66.7	305	3	US-09-567-969-22	Sequence 22, Appl
36	32	66.7	305	3	US-09-568-480-22	Sequence 22, Appl
37	32	66.7	305	3	US-09-568-486-22	Sequence 22, Appl
38	32	66.7	305	3	US-09-568-472-22	Sequence 22, Appl
39	32	66.7	305	3	US-09-567-899-22	Sequence 22, Appl
40	32	66.7	307	4	US-09-328-352-4994	Sequence 4994, Ap
41	32	66.7	315	4	US-09-902-540-11852	Sequence 11852, A
42	32	66.7	383	4	US-09-489-039A-12482	Sequence 12482, A
43	32	66.7	425	2	US-08-986-963-2	Sequence 2, Appl
44	32	66.7	429	4	US-09-583-110-4561	Sequence 4561, Ap
45	32	66.7	430	3	US-09-007-484-2	Sequence 2, Appl
46	32	66.7	430	3	US-09-309-682-2	Sequence 2, Appl
47	32	66.7	430	3	US-09-107-433-2686	Sequence 2686, Ap
48	32	66.7	431	4	US-09-540-236-2280	Sequence 2280, Ap
49	32	66.7	492	3	US-09-413-814-72	Sequence 72, Appl
50	32	66.7	536	4	US-09-328-352-7300	Sequence 7300, Ap
51	32	66.7	536	4	US-09-653-274-10	Sequence 10, Appl
52	32	66.7	536	4	US-10-461-791-10	Sequence 10, Appl
53	32	66.7	930	4	US-08-254-594-6	Sequence 6, Appl
54	32	66.7	1385	2	US-08-687-339-7	Sequence 7, Appl
55	32	66.7	2042	4	US-09-489-039A-14053	Sequence 14053, A
56	31	64.6	82	4	US-09-134-000C-5571	Sequence 5571, Ap
57	31	64.6	129	4	US-09-270-767-32657	Sequence 32657, A
58	31	64.6	129	4	US-09-270-767-47874	Sequence 47874, A
59	31	64.6	174	4	US-09-496-118B-5	Sequence 5, Appl
60	31	64.6	262	4	US-09-252-991A-25231	Sequence 25231, A
61	31	64.6	270	4	US-09-270-767-37404	Sequence 37404, A
62	31	64.6	270	4	US-09-270-767-52621	Sequence 52621, A
63	31	64.6	285	3	US-09-286-529-1	Sequence 1, Appl
64	31	64.6	285	4	US-09-589-287B-2	Sequence 2, Appl
65	31	64.6	285	4	US-09-496-118B-1	Sequence 1, Appl
66	31	64.6	285	4	US-09-565-423-2	Sequence 2, Appl
67	31	64.6	285	4	US-09-879-919-23	Sequence 23, Appl
68	31	64.6	285	4	US-09-588-947A-2	Sequence 2, Appl
69	31	64.6	285	4	US-09-589-286A-2	Sequence 2, Appl
70	31	64.6	285	4	US-09-005-874-2	Sequence 2, Appl
71	31	64.6	285	4	US-09-255-794A-2	Sequence 2, Appl
72	31	64.6	285	4	US-09-507-968D-2	Sequence 2, Appl
73	31	64.6	299	4	US-09-949-016-8108	Sequence 8108, Ap
74	31	64.6	315	4	US-09-248-796A-19855	Sequence 19855, A
75	31	64.6	328	3	US-08-858-003-31	Sequence 31, Appl
76	31	64.6	328	3	US-09-078-166-31	Sequence 31, Appl
77	31	64.6	328	3	US-08-997-467-31	Sequence 31, Appl
78	31	64.6	364	4	US-09-438-046-13	Sequence 13, Appl
79	31	64.6	417	4	US-09-533-494A-2	Sequence 2, Appl
80	31	64.6	566	4	US-09-514-245-21	Sequence 21, Appl
81	31	64.6	668	1	US-08-530-950-13	Sequence 13, Appl
82	31	64.6	668	3	US-09-149-879-13	Sequence 13, Appl
83	31	64.6	668	4	US-09-057-009-13	Sequence 13, Appl
84	31	64.6	732	4	US-09-252-991A-32240	Sequence 32240, A
85	31	64.6	775	4	US-10-148-806-37	Sequence 37, Appl
86	31	64.6	1040	2	US-08-254-989-2	Sequence 2, Appl
87	31	64.6	1043	4	US-09-538-052-935	Sequence 935, App
88	31	64.6	1302	4	US-09-902-540-14853	Sequence 14853, A
89	31	64.6	1709	4	US-09-949-016-10503	Sequence 10503, A
90	31	64.6	1770	4	US-09-902-540-10224	Sequence 10224, A
91	31	64.6	2594	3	US-08-718-388-7	Sequence 7, Appl
92	31	64.6	5405	3	US-08-718-388-9	Sequence 9, Appl
93	31	64.6	15281	2	US-08-471-119A-2	Sequence 2, Appl
94	30	62.5	25	1	US-07-792-259-3	Sequence 3, Appl
95	30	62.5	42	2	US-08-441-871-82	Sequence 82, Appl
96	30	62.5	62	4	US-09-107-532A-5561	Sequence 5561, Ap
97	30	62.5	71	4	US-09-248-796A-21236	Sequence 21236, A
98	30	62.5	74	4	US-09-248-796A-25684	Sequence 25684, A
99	30	62.5	158	2	US-08-403-852D-29	Sequence 29, Appl
100	30	62.5	158	3	US-08-510-646B-30	Sequence 30, Appl

101	30	62.5	158	3	US-09-231-818-29	Sequence 29, Appl	174	29	60.4	395	4	US-09-710-279-26	Sequence 26, Appl
102	30	62.5	158	4	US-09-635-359B-29	Sequence 29, Appl	175	29	60.4	402	3	US-09-252-991A-31080	Sequence 31080, A
103	30	62.5	179	4	US-09-615-192A-289	Sequence 289, App	176	29	60.4	406	3	US-09-134-000C-5330	Sequence 5330, Ap
104	30	62.5	202	4	US-09-134-000C-3978	Sequence 3978, App	177	29	60.4	416	4	US-09-134-000C-5375	Sequence 5375, Ap
105	30	62.5	221	4	US-09-248-796A-16638	Sequence 16638, A	178	29	60.4	422	4	US-09-252-991A-21999	Sequence 21999, A
106	30	62.5	223	4	US-09-328-452-4358	Sequence 4358, Ap	179	29	60.4	426	4	US-09-711-164-380	Sequence 380, App
107	30	62.5	264	4	US-09-543-681A-6332	Sequence 6332, Ap	180	29	60.4	470	4	US-09-270-767-61004	Sequence 61004, A
108	30	62.5	266	4	US-09-134-000C-3456	Sequence 3456, Ap	181	29	60.4	485	4	US-09-252-991A-27501	Sequence 27501, A
109	30	62.5	304	4	US-09-543-681A-8200	Sequence 8200, Ap	182	29	60.4	513	4	US-09-142-108C-42	Sequence 42, Appl
110	30	62.5	331	4	US-09-902-540-10221	Sequence 10221, A	183	29	60.4	518	2	US-08-836-620A-18	Sequence 18, Appl
111	30	62.5	339	4	US-09-489-039A-12231	Sequence 12231, A	184	29	60.4	568	1	US-07-803-622B-4	Sequence 4, Appl
112	30	62.5	343	2	US-08-624-545-1	Sequence 1, Appli	185	29	60.4	582	3	US-09-242-632A-2	Sequence 2, Appli
113	30	62.5	343	2	US-09-235-736-1	Sequence 1, Appli	186	29	60.4	582	3	US-09-242-632A-2	Sequence 2, Appli
114	30	62.5	344	3	US-09-024-532-4	Sequence 4, Appli	187	29	60.4	601	4	US-09-270-767-45496	Sequence 45496, A
115	30	62.5	344	3	US-09-705-185-4	Sequence 4, Appli	188	29	60.4	615	4	US-09-198-452A-1037	Sequence 1037, Ap
116	30	62.5	345	1	US-07-792-259-12	Sequence 12, Appl	189	29	60.4	668	4	US-09-252-991A-22794	Sequence 22794, A
117	30	62.5	363	1	US-08-458-023B-6	Sequence 6, Appli	190	29	60.4	677	2	US-08-522-269B-3	Sequence 3, Appli
118	30	62.5	364	1	US-07-792-259-17	Sequence 17, Appl	191	29	60.4	777	4	US-09-294-923-3	Sequence 2, Appli
119	30	62.5	368	4	US-09-134-000C-4546	Sequence 4546, Ap	192	29	60.4	753	3	US-09-276-400-2	Sequence 2, Appli
120	30	62.5	384	4	US-09-252-991A-26093	Sequence 26093, A	193	29	60.4	753	3	US-09-448-076-2	Sequence 2, Appli
121	30	62.5	385	4	US-09-252-991A-26432	Sequence 26432, A	194	29	60.4	753	3	US-09-702-572-2	Sequence 2, Appli
122	30	62.5	392	2	US-08-387-942C-23	Sequence 23, Appl	195	29	60.4	768	4	US-09-134-000C-6467	Sequence 6467, Ap
123	30	62.5	435	2	US-08-331-515A-2	Sequence 2, Appli	196	29	60.4	777	4	US-09-811-469-2	Sequence 2, Appli
124	30	62.5	435	3	US-09-168-406A-2	Sequence 2, Appli	197	29	60.4	777	4	US-10-370-659-2	Sequence 2, Appli
125	30	62.5	446	4	US-09-949-016-7571	Sequence 7571, Ap	198	29	60.4	885	4	US-09-502-540-13431	Sequence 13431, A
126	30	62.5	480	4	US-09-252-991A-20125	Sequence 20125, A	199	29	60.4	929	4	US-09-254-594-3	Sequence 3, Appli
127	30	62.5	493	4	US-09-583-110-2996	Sequence 2996, Ap	200	29	60.4	981	4	US-09-991-258-13	Sequence 13, Appl
128	30	62.5	495	4	US-09-107-532A-6687	Sequence 6687, Ap	201	29	60.4	998	4	US-09-198-452A-841	Sequence 841, App
129	30	62.5	499	4	US-09-107-433-4478	Sequence 4478, Ap	202	29	60.4	998	4	US-09-438-185A-793	Sequence 793, App
130	30	62.5	502	4	US-09-134-000C-6237	Sequence 6237, Ap	203	29	60.4	1306	3	US-08-999-774A-13	Sequence 13, Appl
131	30	62.5	502	4	US-09-949-016-5991	Sequence 5991, Ap	204	29	60.4	1317	4	US-09-949-016-7588	Sequence 7588, Ap
132	30	62.5	513	4	US-09-465-558-36	Sequence 36, Appl	205	29	60.4	1326	4	US-09-688-188B-15	Sequence 15, Appl
133	30	62.5	553	2	US-08-387-942C-2	Sequence 2, Appli	206	29	60.4	1326	4	US-09-291-417D-15	Sequence 15, Appl
134	30	62.5	649	4	US-09-902-540-13928	Sequence 13928, A	207	29	60.4	1805	4	US-09-949-016-8246	Sequence 8246, Ap
135	30	62.5	769	4	US-09-949-016-10665	Sequence 10665, A	208	29	60.4	2482	1	US-08-328-254-6	Sequence 6, Appli
136	30	62.5	923	4	US-09-538-092-80	Sequence 80, Appl	209	29	60.4	3210	4	US-09-538-092-1154	Sequence 1154, Ap
137	30	62.5	1437	4	US-09-949-016-10535	Sequence 10535, A	210	29	60.4	3218	1	US-08-764-100-27	Sequence 27, Appl
138	29	60.4	11	4	US-09-082-358B-87	Sequence 87, Appl	211	29	60.4	3248	5	US-08-353-700-1	Sequence 1, Appli
139	29	60.4	61	4	US-09-107-532A-4643	Sequence 4643, Ap	212	28.5	59.4	3248	4	PCT-US95-16216-1	Sequence 2, Appli
140	29	60.4	71	4	US-09-134-000C-6796	Sequence 6796, Ap	213	28.5	59.4	180	4	US-09-465-558-2	Sequence 2, Appli
141	29	60.4	138	4	US-09-252-991A-21153	Sequence 21153, A	214	28.5	59.4	241	4	US-09-270-767-34316	Sequence 34316, A
142	29	60.4	140	4	US-09-134-000C-6375	Sequence 6375, Ap	215	28.5	59.4	241	4	US-09-270-767-49533	Sequence 49533, A
143	29	60.4	142	4	US-09-902-540-12459	Sequence 12459, A	216	28.5	59.4	510	4	US-09-465-558-40	Sequence 40, Appl
144	29	60.4	150	4	US-09-142-108C-10	Sequence 10, Appl	217	28	58.3	48	3	US-09-346-860-10	Sequence 10, Appl
145	29	60.4	166	4	US-09-270-767-49327	Sequence 49327, A	218	28	58.3	48	3	US-09-735-685-10	Sequence 10, Appl
146	29	60.4	183	4	US-09-270-767-38148	Sequence 38148, A	219	28	58.3	60	4	US-09-621-976-5575	Sequence 5575, Ap
147	29	60.4	183	4	US-09-270-767-53365	Sequence 53365, A	220	28	58.3	64	4	US-09-149-476-330	Sequence 330, App
148	29	60.4	186	4	US-09-270-767-35570	Sequence 35570, A	221	28	58.3	64	4	US-09-959-897-37	Sequence 37, Appl
149	29	60.4	186	4	US-09-270-767-50787	Sequence 50787, A	222	28	58.3	72	4	US-09-270-767-58180	Sequence 58180, A
150	29	60.4	192	4	US-09-543-681A-8139	Sequence 8139, Ap	223	28	58.3	72	4	US-09-248-796A-26744	Sequence 26744, A
151	29	60.4	193	4	US-09-270-767-3468	Sequence 3468, A	224	28	58.3	76	4	US-09-902-540-15288	Sequence 15288, A
152	29	60.4	204	4	US-09-270-767-32818	Sequence 32818, A	225	28	58.3	96	4	US-09-673-395A-418	Sequence 418, App
153	29	60.4	204	4	US-09-270-767-48035	Sequence 48035, A	226	28	58.3	96	4	US-09-673-395A-463	Sequence 463, App
154	29	60.4	228	4	US-09-248-796A-17522	Sequence 17522, A	227	28	58.3	107	4	US-09-270-767-59736	Sequence 59736, A
155	29	60.4	252	4	US-09-489-039A-12324	Sequence 12324, A	228	28	58.3	108	4	US-09-270-767-31878	Sequence 31878, A
156	29	60.4	254	4	US-09-978-758-2	Sequence 2, Appli	229	28	58.3	108	4	US-09-270-767-44313	Sequence 44313, A
157	29	60.4	257	4	US-09-540-236-2812	Sequence 2812, Ap	230	28	58.3	108	4	US-09-270-767-47095	Sequence 47095, A
158	29	60.4	284	4	US-09-489-039A-9889	Sequence 9889, Ap	231	28	58.3	114	3	US-09-134-001C-4120	Sequence 4120, Ap
159	29	60.4	296	4	US-09-540-236-2650	Sequence 2650, Ap	232	28	58.3	116	1	US-08-122-546-8	Sequence 8, Appli
160	29	60.4	311	4	US-09-270-767-33374	Sequence 33374, A	233	28	58.3	116	3	US-09-131-053A-8	Sequence 8, Appli
161	29	60.4	311	4	US-09-270-767-48591	Sequence 48591, A	234	28	58.3	116	3	US-09-131-053A-8	Sequence 8, Appli
162	29	60.4	313	4	US-09-248-796A-17621	Sequence 17621, A	235	28	58.3	123	4	US-09-489-039A-11787	Sequence 11787, A
163	29	60.4	318	4	US-09-107-532A-5738	Sequence 5738, Ap	236	28	58.3	137	4	US-09-252-991A-22501	Sequence 22501, A
164	29	60.4	329	4	US-08-981-527A-8	Sequence 8, Appli	237	28	58.3	150	4	US-09-270-767-32176	Sequence 32176, A
165	29	60.4	329	4	US-10-147-936B-8	Sequence 8, Appli	238	28	58.3	150	4	US-09-270-767-47393	Sequence 47393, A
166	29	60.4	344	3	US-09-393-554-5	Sequence 5, Appli	239	28	58.3	151	1	US-08-318-947A-9	Sequence 9, Appli
167	29	60.4	346	4	US-09-248-796A-25822	Sequence 25822, A	240	28	58.3	151	2	US-08-795-303-9	Sequence 9, Appli
168	29	60.4	350	4	US-09-48-796A-17746	Sequence 17746, A	241	28	58.3	155	4	US-09-270-767-40394	Sequence 40394, A
169	29	60.4	377	1	US-08-153-848-42	Sequence 42, Appl	242	28	58.3	155	4	US-09-270-767-55610	Sequence 55610, A
170	29	60.4	377	3	US-09-299-843A-42	Sequence 42, Appl	243	28	58.3	156	4	US-09-902-540-14827	Sequence 14827, A
171	29	60.4	377	3	US-09-088-337B-42	Sequence 42, Appl	244	28	58.3	201	4	US-09-107-433-2650	Sequence 2650, Ap
172	29	60.4	377	3	PCT-US93-11153-42	Sequence 42, Appl	245	28	58.3	201	4	US-09-902-540-12974	Sequence 12974, A
173	29	60.4	377	5		Sequence 42, Appl	246	28	58.3	204	4	US-09-583-110-4824	Sequence 4824, Ap



247	28	58.3	208	4	US-09-673-395A-516	Sequence 516, App	320	28	58.3	855	4	US-09-107-532A-5646	Sequence 5646, Ap
248	28	58.3	220	4	US-09-252-991A-21239	Sequence 21239, A	321	28	58.3	858	4	US-09-252-991A-23178	Sequence 23178, A
249	28	58.3	226	4	US-09-328-352-5984	Sequence 5984, Ap	322	28	58.3	890	1	US-08-472-934-8	Sequence 8, Appli
250	28	58.3	237	4	US-09-583-110-5199	Sequence 5199, Ap	323	28	58.3	890	2	US-08-323-460A-8	Sequence 8, Appli
251	28	58.3	241	4	US-09-107-433-3412	Sequence 3412, Ap	324	28	58.3	890	2	US-08-461-146C-8	Sequence 8, Appli
252	28	58.3	244	4	US-09-902-540-16220	Sequence 16220, A	325	28	58.3	890	3	US-08-461-145C-8	Sequence 12, Appli
253	28	58.3	254	4	US-09-252-991A-25576	Sequence 25576, A	326	28	58.3	890	3	US-08-628-829-12	Sequence 12, Appli
254	28	58.3	264	4	US-09-270-767-33544	Sequence 33544, A	327	28	58.3	910	4	US-09-949-016-7055	Sequence 7055, Ap
255	28	58.3	264	4	US-09-270-767-48761	Sequence 48761, A	328	28	58.3	977	4	US-09-590-656-1	Sequence 1, Appli
256	28	58.3	265	3	US-09-154-083-8	Sequence 8, Appli	329	28	58.3	977	4	US-09-733-764-1	Sequence 1, Appli
257	28	58.3	268	4	US-09-252-991A-21630	Sequence 21630, A	330	28	58.3	985	2	US-08-867-941-13	Sequence 13, Appli
258	28	58.3	272	4	US-09-949-016-6331	Sequence 6331, Ap	331	28	58.3	985	2	US-08-867-941-17	Sequence 17, Appli
259	28	58.3	296	4	US-09-270-767-42861	Sequence 42861, A	332	28	58.3	985	3	US-09-074-658-13	Sequence 13, Appli
260	28	58.3	300	4	US-09-252-991A-23342	Sequence 23342, A	333	28	58.3	985	3	US-09-074-658-17	Sequence 17, Appli
261	28	58.3	331	4	US-09-546-986A-4	Sequence 4, Appli	334	28	58.3	1000	2	US-08-867-941-12	Sequence 12, Appli
262	28	58.3	331	4	US-09-524-730-4	Sequence 730, A	335	28	58.3	1000	2	US-08-867-941-16	Sequence 16, Appli
263	28	58.3	332	4	US-09-949-016-10159	Sequence 10159, A	336	28	58.3	1000	3	US-09-074-658-12	Sequence 12, Appli
264	28	58.3	341	4	US-09-252-991A-22453	Sequence 22453, A	337	28	58.3	1000	3	US-09-074-658-16	Sequence 16, Appli
265	28	58.3	342	4	US-09-248-796A-19527	Sequence 19527, A	338	28	58.3	1003	4	US-09-540-236-2757	Sequence 2757, Ap
266	28	58.3	344	4	US-09-252-991A-27871	Sequence 27871, A	339	28	58.3	1101	2	US-08-469-537A-96	Sequence 96, Appli
267	28	58.3	375	4	US-09-902-540-12456	Sequence 12456, A	340	28	58.3	1118	1	US-07-934-393B-2	Sequence 2, Appli
268	28	58.3	392	4	US-09-071-035-260	Sequence 260, App	341	28	58.3	1118	1	US-08-838-957A-2	Sequence 2, Appli
269	28	58.3	394	4	US-09-949-016-6716	Sequence 6716, Ap	342	28	58.3	1118	2	US-08-278-089A-6	Sequence 6, Appli
270	28	58.3	404	1	US-08-696-770-2	Sequence 2, Appli	343	28	58.3	1122	2	US-08-838-957A-6	Sequence 6, Appli
271	28	58.3	404	2	US-09-015-557-2	Sequence 2, Appli	344	28	58.3	1122	2	US-08-323-474-2	Sequence 2, Appli
272	28	58.3	408	4	US-09-583-110-5062	Sequence 5062, Ap	345	28	58.3	1124	1	US-09-949-016-5946	Sequence 5946, Ap
273	28	58.3	410	4	US-09-489-039A-7598	Sequence 7598, Ap	346	28	58.3	1124	5	PCT-US93-06093-2	Sequence 2, Appli
274	28	58.3	416	3	US-08-978-289-12	Sequence 12, Appli	347	28	58.3	1124	5	PCT-US93-06093-2	Sequence 2, Appli
275	28	58.3	416	4	US-09-601-478-1	Sequence 1, Appli	348	28	58.3	1157	4	US-09-949-016-9568	Sequence 9568, Ap
276	28	58.3	416	4	US-09-601-478-4	Sequence 4, Appli	349	28	58.3	1157	4	US-09-583-110-3634	Sequence 3634, Ap
277	28	58.3	417	4	US-09-252-991A-20165	Sequence 20165, A	350	28	58.3	1203	4	US-09-107-433-4255	Sequence 4255, Ap
278	28	58.3	418	4	US-09-489-039A-11946	Sequence 11946, A	351	28	58.3	1203	4	US-09-107-433-4255	Sequence 10, Appli
279	28	58.3	423	4	US-09-107-433-2664	Sequence 2664, Ap	352	28	58.3	1247	2	US-08-323-460A-10	Sequence 10, Appli
280	28	58.3	424	4	US-09-796-487-8	Sequence 8, Appli	353	28	58.3	1247	2	US-08-461-146C-10	Sequence 10, Appli
281	28	58.3	427	4	US-09-949-016-9842	Sequence 9842, Ap	354	28	58.3	1247	3	US-08-461-145C-10	Sequence 10, Appli
282	28	58.3	431	4	US-09-543-681A-6326	Sequence 6326, Ap	355	28	58.3	1275	4	US-09-902-540-10769	Sequence 10769, A
283	28	58.3	437	4	US-09-252-991A-18636	Sequence 18636, Ap	356	28	58.3	1279	4	US-09-170-496D-293	Sequence 293, App
284	28	58.3	440	4	US-09-252-991A-26855	Sequence 26855, A	357	28	58.3	1279	4	US-09-364-425B-58	Sequence 58, Appli
285	28	58.3	440	4	US-09-489-039A-12067	Sequence 12067, A	358	28	58.3	1326	4	US-09-489-039A-7584	Sequence 7584, Ap
286	28	58.3	443	4	US-09-107-532A-6063	Sequence 6063, Ap	359	28	58.3	1330	4	US-09-543-681A-8057	Sequence 8057, Ap
287	28	58.3	448	4	US-09-361-443-4	Sequence 4, Appli	360	28	58.3	1339	3	US-08-462-467B-14	Sequence 14, Appli
288	28	58.3	461	4	US-09-252-991A-26351	Sequence 26351, A	361	28	58.3	1431	4	US-09-902-540-10614	Sequence 10614, A
289	28	58.3	467	4	US-09-361-443-2	Sequence 2, Appli	362	28	58.3	1462	4	US-09-538-092-1043	Sequence 1043, Ap
290	28	58.3	469	4	US-09-198-452A-793	Sequence 793, App	363	28	58.3	1495	3	US-08-462-467B-12	Sequence 12, Appli
291	28	58.3	469	4	US-09-438-185A-745	Sequence 745, App	364	28	58.3	1514	2	US-08-853-310-4	Sequence 4, Appli
292	28	58.3	470	4	US-09-543-681A-5952	Sequence 5952, Ap	365	28	58.3	1597	3	US-09-423-890-13	Sequence 13, Appli
293	28	58.3	473	4	US-09-949-016-9481	Sequence 9481, Ap	366	28	58.3	1597	3	US-08-628-829-14	Sequence 14, Appli
294	28	58.3	517	4	US-09-142-108C-21	Sequence 21, Appli	367	28	58.3	1638	4	US-09-071-035-258	Sequence 258, App
295	28	58.3	543	4	US-09-248-796A-14702	Sequence 14702, A	368	28	58.3	1638	4	US-09-071-035-282	Sequence 282, App
296	28	58.3	553	1	US-08-565-386-12	Sequence 12, Appli	369	28	58.3	1638	4	US-09-071-035-282	Sequence 266, App
297	28	58.3	574	4	US-09-252-991A-30868	Sequence 30868, A	370	28	58.3	1747	4	US-09-134-000C-5999	Sequence 5999, Ap
298	28	58.3	579	4	US-09-543-681A-5050	Sequence 5050, Ap	371	28	58.3	1996	2	US-08-804-227C-9	Sequence 9, Appli
299	28	58.3	585	2	US-08-867-941-21	Sequence 21, Appli	372	28	58.3	1996	2	US-08-804-198-3	Sequence 3, Appli
300	28	58.3	585	3	US-09-074-658-21	Sequence 21, Appli	373	28	58.3	2432	3	US-09-074-658-15	Sequence 15, Appli
301	28	58.3	596	3	US-09-171-337A-6	Sequence 6, Appli	374	28	58.3	2432	3	US-09-074-658-11	Sequence 11, Appli
302	28	58.3	596	4	US-09-631-022-6	Sequence 6, Appli	375	28	58.3	2451	3	US-09-949-016-9675	Sequence 15, Appli
303	28	58.3	596	4	US-09-134-000C-4516	Sequence 4516, Ap	376	28	58.3	2787	3	US-09-245-041-15	Sequence 15, Appli
304	28	58.3	651	4	US-09-902-540-11066	Sequence 11066, A	377	28	58.3	2787	4	US-09-358-055B-15	Sequence 15, Appli
305	28	58.3	667	4	US-09-710-279-198	Sequence 198, App	378	28	58.3	2787	4	US-09-893-238-15	Sequence 15, Appli
306	28	58.3	669	4	US-09-071-035-264	Sequence 264, App	379	28	58.3	3562	4	US-09-679-279-14	Sequence 14, Appli
307	28	58.3	681	3	US-08-760-615-6	Sequence 6, Appli	380	28	58.3	3562	2	US-07-642-734C-4	Sequence 4, Appli
308	28	58.3	684	4	US-09-252-991A-24802	Sequence 24802, A	381	28	58.3	3567	3	US-08-439-009A-4	Sequence 4, Appli
309	28	58.3	704	4	US-09-590-656-2	Sequence 2, Appli	382	28	58.3	3567	3	US-09-949-016-7876	Sequence 7876, Ap
310	28	58.3	711	3	US-08-949-588-2	Sequence 2, Appli	383	28	58.3	3854	2	US-08-804-227C-8	Sequence 8, Appli
311	28	58.3	723	3	US-09-134-001C-5060	Sequence 5060, Ap	384	28	58.3	4550	2	US-08-804-198-2	Sequence 2, Appli
312	28	58.3	749	4	US-09-562-737-97	Sequence 97, Appli	385	28	58.3	4550	2	US-09-902-540-16789	Sequence 14362, A
313	28	58.3	753	2	US-08-867-941-20	Sequence 20, Appli	386	27.5	57.3	239	4	US-09-902-540-16789	Sequence 16, Appli
314	28	58.3	753	3	US-09-074-658-20	Sequence 20, Appli	387	27.5	57.3	293	4	US-09-292-097-16	Sequence 16, Appli
315	28	58.3	792	3	US-09-543-681A-6617	Sequence 6617, Ap	388	27.5	57.3	402	3	US-09-933-561-16	Sequence 16, Appli
316	28	58.3	796	3	US-09-005-180A-1	Sequence 1, Appli	389	27.5	57.3	476	4	US-09-248-796A-17464	Sequence 17464, A
317	28	58.3	820	4	US-09-902-540-16384	Sequence 16384, A	390	27.5	57.3	1558	4	US-09-949-016-6803	Sequence 6803, Ap
318	28	58.3	832	4	US-09-252-991A-24866	Sequence 24866, A	391	27.5	57.3	1606	4	US-09-949-016-7371	Sequence 7371, Ap
319	28	58.3					392	27.5	57.3				

393	27	56.2	14	3	US-09-306-756-2	Sequence 2, Appli	466	27	56.2	300	2	US-08-794-796-2	Sequence 2, Appli
394	27	56.2	14	4	US-09-674-307-2	Sequence 2, Appli	467	27	56.2	300	4	US-09-585-858-45	Sequence 45, Appli
395	27	56.2	62	3	US-08-936-165A-364	Sequence 364, App	468	27	56.2	300	4	US-09-632-277A-2	Sequence 2, Appli
396	27	56.2	64	4	US-09-248-796A-27846	Sequence 27846, A	469	27	56.2	300	4	US-09-523-323-52	Sequence 52, Appli
397	27	56.2	79	3	US-09-270-767-58435	Sequence 58435, A	470	27	56.2	300	4	US-09-896-096A-1	Sequence 1, Appli
398	27	56.2	79	3	US-09-100-802-4	Sequence 4, Appli	471	27	56.2	300	4	US-10-270-878-45	Sequence 45, Appli
399	27	56.2	83	4	US-09-248-796A-26718	Sequence 26718, A	472	27	56.2	300	4	US-09-936-019-3	Sequence 3, Appli
400	27	56.2	85	4	US-09-270-767-61052	Sequence 61052, A	473	27	56.2	301	3	US-09-002-298-3	Sequence 3, Appli
401	27	56.2	88	2	US-08-465-380-43	Sequence 43, Appli	474	27	56.2	301	4	US-09-481-277-3	Sequence 3, Appli
402	27	56.2	88	2	US-08-486-397-43	Sequence 43, Appli	475	27	56.2	301	4	US-09-252-991A-27392	Sequence 27392, A
403	27	56.2	88	2	US-08-486-399-43	Sequence 43, Appli	476	27	56.2	301	4	US-08-311-931A-308	Sequence 308, App
404	27	56.2	88	2	US-08-461-965-43	Sequence 43, Appli	477	27	56.2	301	4	US-09-270-767-35774	Sequence 35774, A
405	27	56.2	88	2	US-08-634-641-43	Sequence 43, Appli	478	27	56.2	301	4	US-09-270-767-50991	Sequence 50991, A
406	27	56.2	88	3	US-09-249-471-43	Sequence 43, Appli	479	27	56.2	309	4	US-09-415-260A-1	Sequence 1, Appli
407	27	56.2	88	3	US-09-249-472-43	Sequence 43, Appli	480	27	56.2	309	4	US-09-415-260A-2	Sequence 2, Appli
408	27	56.2	88	3	US-09-249-451-43	Sequence 43, Appli	481	27	56.2	309	4	US-09-415-260A-3	Sequence 3, Appli
409	27	56.2	88	3	US-08-809-455-43	Sequence 43, Appli	482	27	56.2	309	4	US-09-415-260A-4	Sequence 4, Appli
410	27	56.2	88	3	US-09-249-461-43	Sequence 43, Appli	483	27	56.2	309	4	US-09-415-260A-5	Sequence 5, Appli
411	27	56.2	88	3	US-09-249-448-43	Sequence 43, Appli	484	27	56.2	309	4	US-09-415-260A-6	Sequence 6, Appli
412	27	56.2	88	4	US-09-249-473-43	Sequence 43, Appli	485	27	56.2	309	4	US-09-415-260A-7	Sequence 7, Appli
413	27	56.2	97	4	US-09-270-767-46651	Sequence 46651, A	486	27	56.2	309	4	US-09-415-260A-8	Sequence 8, Appli
414	27	56.2	102	4	US-09-902-540-14150	Sequence 14150, A	487	27	56.2	313	4	US-09-602-777A-190	Sequence 190, App
415	27	56.2	105	4	US-09-540-236-2606	Sequence 2606, Ap	488	27	56.2	313	4	US-09-602-777A-440	Sequence 440, App
416	27	56.2	111	4	US-09-270-767-33866	Sequence 33866, A	489	27	56.2	315	4	US-09-543-681A-7121	Sequence 7121, Ap
417	27	56.2	111	4	US-09-270-767-49083	Sequence 49083, A	490	27	56.2	315	4	US-09-270-767-33342	Sequence 33342, A
418	27	56.2	118	3	US-08-301-162-10	Sequence 10, Appli	491	27	56.2	315	4	US-09-270-767-48559	Sequence 48559, A
419	27	56.2	118	3	US-08-461-240-10	Sequence 10, Appli	492	27	56.2	325	1	US-07-626-589-2	Sequence 2, Appli
420	27	56.2	118	4	US-09-968-927-10	Sequence 10, Appli	493	27	56.2	325	1	US-08-190-082A-57	Sequence 57, Appli
421	27	56.2	126	1	US-08-044-547-3	Sequence 3, Appli	494	27	56.2	325	1	US-08-323-444A-1	Sequence 1, Appli
422	27	56.2	129	4	US-09-107-532A-6426	Sequence 6426, Ap	495	27	56.2	325	1	US-08-236-410-2	Sequence 2, Appli
423	27	56.2	130	4	US-09-902-540-11863	Sequence 11863, A	496	27	56.2	325	1	US-08-465-421-2	Sequence 2, Appli
424	27	56.2	139	4	US-09-270-767-38418	Sequence 38418, A	497	27	56.2	325	3	US-08-477-346-57	Sequence 57, Appli
425	27	56.2	139	4	US-09-270-767-53635	Sequence 53635, A	498	27	56.2	325	3	US-08-473-089-57	Sequence 57, Appli
426	27	56.2	141	4	US-09-489-039A-9720	Sequence 9720, Ap	499	27	56.2	325	4	US-08-487-072A-57	Sequence 57, Appli
427	27	56.2	145	4	US-09-732-210-1321	Sequence 1321, Ap	500	27	56.2	326	4	US-09-198-452A-150	Sequence 150, App
428	27	56.2	146	4	US-09-893-737-88	Sequence 88, Appli	501	27	56.2	326	4	US-09-438-185A-134	Sequence 134, App
429	27	56.2	146	4	US-09-523-323-59	Sequence 59, Appli	502	27	56.2	326	4	US-09-902-540-11608	Sequence 11608, A
430	27	56.2	147	1	US-08-044-547-1	Sequence 1, Appli	503	27	56.2	330	4	US-09-107-532A-7050	Sequence 7050, Ap
431	27	56.2	149	4	US-09-270-767-47885	Sequence 47885, A	504	27	56.2	333	4	US-09-949-016-7678	Sequence 7678, Ap
432	27	56.2	153	3	US-09-286-529-2	Sequence 2, Appli	505	27	56.2	338	4	US-09-252-991A-26745	Sequence 26745, A
433	27	56.2	153	4	US-09-270-767-3444	Sequence 3444, A	506	27	56.2	338	4	US-09-949-016-7050	Sequence 7050, Ap
434	27	56.2	153	4	US-09-270-767-49661	Sequence 49661, A	507	27	56.2	341	4	US-09-902-540-13885	Sequence 13885, A
435	27	56.2	156	4	US-09-270-767-32472	Sequence 32472, A	508	27	56.2	346	4	US-09-252-991A-31465	Sequence 31465, A
436	27	56.2	156	4	US-09-270-767-47689	Sequence 47689, A	509	27	56.2	348	4	US-09-489-039A-8827	Sequence 8827, Ap
437	27	56.2	162	4	US-09-328-352-5894	Sequence 5894, Ap	510	27	56.2	351	4	US-09-902-540-12944	Sequence 12944, A
438	27	56.2	163	4	US-09-465-558-8	Sequence 8, Appli	511	27	56.2	352	4	US-09-252-991A-19989	Sequence 19989, A
439	27	56.2	170	3	US-09-189-035-2	Sequence 2, Appli	512	27	56.2	352	4	US-09-711-164-448	Sequence 448, App
440	27	56.2	170	3	US-09-382-086-2	Sequence 2, Appli	513	27	56.2	353	4	US-09-902-540-14186	Sequence 14186, A
441	27	56.2	171	4	US-09-489-039A-8636	Sequence 8636, Ap	514	27	56.2	356	3	US-09-235-103-4	Sequence 4, Appli
442	27	56.2	173	4	US-09-252-991A-25090	Sequence 25090, A	515	27	56.2	356	3	US-09-235-103-13	Sequence 13, Appli
443	27	56.2	182	4	US-10-085-572-8	Sequence 8, Appli	516	27	56.2	363	4	US-09-270-767-45543	Sequence 45543, A
444	27	56.2	182	4	US-10-087-195-8	Sequence 8, Appli	517	27	56.2	363	4	US-09-949-016-9869	Sequence 9869, Ap
445	27	56.2	184	4	US-09-270-767-43100	Sequence 43100, A	518	27	56.2	364	4	US-09-270-767-42122	Sequence 42122, A
446	27	56.2	191	4	US-09-538-092-140	Sequence 140, App	519	27	56.2	368	4	US-09-614-912-38	Sequence 38, Appli
447	27	56.2	193	4	US-09-902-540-11304	Sequence 11304, A	520	27	56.2	368	4	US-09-248-796A-20840	Sequence 20840, A
448	27	56.2	200	4	US-09-618-919A-20	Sequence 20, Appli	521	27	56.2	370	3	US-09-134-001C-3769	Sequence 3769, Ap
449	27	56.2	206	4	US-09-252-991A-17940	Sequence 17940, A	522	27	56.2	372	4	US-09-328-352-6955	Sequence 6955, Ap
450	27	56.2	209	4	US-09-489-039A-13956	Sequence 13956, A	523	27	56.2	376	3	US-09-056-556-202	Sequence 202, App
451	27	56.2	211	3	US-09-286-529-20	Sequence 20, Appli	524	27	56.2	376	4	US-09-072-596-197	Sequence 197, App
452	27	56.2	218	4	US-09-248-796A-15868	Sequence 15868, A	525	27	56.2	376	4	US-09-072-967-202	Sequence 202, App
453	27	56.2	233	2	US-09-151-611-1	Sequence 1, Appli	526	27	56.2	379	4	US-09-491-577-14	Sequence 14, Appli
454	27	56.2	233	3	US-09-370-102-1	Sequence 1, Appli	527	27	56.2	379	4	US-09-602-777A-112	Sequence 112, App
455	27	56.2	245	4	US-08-311-731A-119	Sequence 119, App	528	27	56.2	389	4	US-09-710-279-618	Sequence 618, App
456	27	56.2	249	4	US-09-543-681A-6859	Sequence 6859, Ap	529	27	56.2	391	4	US-09-252-991A-17690	Sequence 17690, A
457	27	56.2	264	2	US-07-857-224B-67	Sequence 67, Appli	530	27	56.2	409	4	US-09-538-092-1214	Sequence 1214, Ap
458	27	56.2	267	4	US-09-248-796A-16355	Sequence 16355, A	531	27	56.2	409	4	US-09-949-016-6281	Sequence 6281, Ap
459	27	56.2	268	2	US-07-857-224B-33	Sequence 33, Appli	532	27	56.2	419	3	US-08-974-691-3	Sequence 3, Appli
460	27	56.2	269	4	US-09-252-991A-29917	Sequence 29917, A	533	27	56.2	419	4	US-09-705-448-10	Sequence 10, Appli
461	27	56.2	270	3	US-09-134-001C-4376	Sequence 4376, Ap	534	27	56.2	429	4	US-09-134-000C-4847	Sequence 4847, Ap
462	27	56.2	271	4	US-09-936-019-1	Sequence 1, Appli	535	27	56.2	430	1	US-08-785-076-3	Sequence 3, Appli
463	27	56.2	272	4	US-09-328-352-7112	Sequence 7112, Ap	536	27	56.2	430	1	US-09-018-824-3	Sequence 3, Appli
464	27	56.2	291	4	US-09-902-540-15646	Sequence 15646, A	537	27	56.2	430	4	US-09-489-039A-10370	Sequence 10370, A
465	27	56.2	299	3	US-09-286-529-17	Sequence 17, Appli	538	27	56.2	432	1	US-08-785-076-2	Sequence 2, Appli

539	27	56.2	432	4	US-09-018-824-2	Sequence 2, Appli	612	27	56.2	2325	3	US-08-695-651-6	Sequence 6, Appli
540	27	56.2	434	4	US-09-198-452A-118	Sequence 118, App	613	27	56.2	2325	3	US-08-930-285-6	Sequence 6, Appli
541	27	56.2	434	4	US-09-438-185A-102	Sequence 102, App	614	27	56.2	2325	3	US-08-695-421-6	Sequence 6, Appli
542	27	56.2	438	3	US-09-134-001C-3139	Sequence 3139, Ap	615	27	56.2	2325	4	US-08-697-826A-10	Sequence 10, Appli
543	27	56.2	442	4	US-09-489-039A-12384	Sequence 12384, A	616	27	56.2	2887	3	US-08-462-467B-2	Sequence 2, Appli
544	27	56.2	446	3	US-09-761-716-2	Sequence 2, Appli	617	27	56.2	2887	3	US-08-462-467B-8	Sequence 8, Appli
545	27	56.2	449	4	US-09-134-000C-4588	Sequence 4588, Ap	618	27	56.2	3433	4	US-09-091-501B-10	Sequence 10, Appli
546	27	56.2	454	4	US-09-949-016-11120	Sequence 11120, A	619	27	56.2	3433	4	US-09-538-092-1136	Sequence 1136, Ap
547	27	56.2	460	4	US-09-252-991A-32040	Sequence 32040, A	620	27	56.2	4019	4	US-09-854-133-425	Sequence 425, App
548	27	56.2	465	4	US-09-902-540-13166	Sequence 13166, A	621	27	56.2	4861	4	US-09-919-497-70	Sequence 70, Appli
549	27	56.2	472	4	US-09-902-540-14925	Sequence 14925, A	622	27	56.2	4866	4	US-09-424-783-3	Sequence 3, Appli
550	27	56.2	475	4	US-09-107-532A-3866	Sequence 3866, Ap	623	27	56.2	4872	4	US-09-424-783-3	Sequence 3, Appli
551	27	56.2	478	4	US-09-252-991A-32557	Sequence 32557, A	624	27	56.2	4968	4	US-09-424-783-5	Sequence 5, Appli
552	27	56.2	489	3	US-09-134-001C-4902	Sequence 4902, Ap	625	27	56.2	5032	4	US-09-538-092-979	Sequence 9, App
553	27	56.2	512	4	US-09-142-108C-6	Sequence 6, Appli	626	27	56.2	5037	4	US-09-424-783-4	Sequence 4, Appli
554	27	56.2	515	4	US-09-453-956-5	Sequence 5, Appli	627	27	56.2	5588	3	US-09-036-987A-6	Sequence 6, Appli
555	27	56.2	519	1	US-08-319-387-4	Sequence 4, Appli	628	27	56.2	5588	3	US-09-370-700-6	Sequence 6, Appli
556	27	56.2	549	4	US-09-543-681A-4720	Sequence 4720, Ap	629	27	56.2	5588	4	US-09-603-207-6	Sequence 6, Appli
557	27	56.2	558	4	US-09-770-517C-2	Sequence 2, Appli	630	26.5	55.2	237	4	US-09-902-540-15928	Sequence 2798, Ap
558	27	56.2	565	4	US-09-489-039A-13004	Sequence 13004, A	631	26.5	55.2	264	4	US-09-107-433-2798	Sequence 42079, A
559	27	56.2	573	4	US-09-134-000C-6366	Sequence 6366, Ap	632	26.5	55.2	405	4	US-09-270-767-42079	Sequence 88, Appli
560	27	56.2	576	4	US-09-578-921A-2	Sequence 2, Appli	633	26.5	55.2	494	3	US-09-413-814-88	Sequence 5, Appli
561	27	56.2	581	4	US-09-489-039A-10603	Sequence 10603, A	634	26.5	55.2	969	4	US-09-321-987B-5	Sequence 5, Appli
562	27	56.2	582	4	US-09-543-681A-5938	Sequence 5938, Ap	635	26.5	55.2	1211	4	US-09-491-522-5	Sequence 5, Appli
563	27	56.2	599	4	US-09-538-092-864	Sequence 864, App	636	26.5	55.2	1211	4	US-09-949-016-11401	Sequence 11401, A
564	27	56.2	605	3	US-09-394-645-2	Sequence 2, Appli	637	26.5	55.2	1276	3	US-08-937-236-3	Sequence 3, Appli
565	27	56.2	605	3	US-09-243-560B-2	Sequence 2, Appli	638	26.5	55.2	1277	3	US-08-937-236-6	Sequence 6, Appli
566	27	56.2	615	4	US-09-345-473E-49	Sequence 49, Appli	639	26.5	55.2	1291	3	US-08-569-214-3	Sequence 3, Appli
567	27	56.2	616	4	US-08-637-670-26	Sequence 26, Appli	640	26.5	55.2	1291	3	US-08-937-236-2	Sequence 2, Appli
568	27	56.2	616	4	US-08-637-670-28	Sequence 28, Appli	641	26.5	55.2	1292	3	US-08-569-214-5	Sequence 5, Appli
569	27	56.2	620	4	US-08-637-670-40	Sequence 40, Appli	642	26.5	55.2	1292	3	US-08-569-214-6	Sequence 6, Appli
570	27	56.2	625	4	US-09-902-540-12827	Sequence 12827, A	643	26.5	55.2	1293	3	US-08-937-236-5	Sequence 5, Appli
571	27	56.2	638	3	US-09-347-488-2	Sequence 2, Appli	644	26.5	55.2	1295	3	US-08-569-214-2	Sequence 2, Appli
572	27	56.2	639	3	US-09-376-856-2	Sequence 2, Appli	645	26	54.2	3892	4	US-09-328-352-5503	Sequence 5503, Ap
573	27	56.2	671	4	US-09-328-352-6003	Sequence 6003, Ap	646	26	54.2	12	3	US-09-181-336-9	Sequence 9, Appli
574	27	56.2	672	4	US-09-336-115C-10	Sequence 10, Appli	647	26	54.2	16	3	US-08-602-999A-221	Sequence 221, App
575	27	56.2	688	4	US-09-949-016-11400	Sequence 11400, A	648	26	54.2	16	4	US-09-500-124-221	Sequence 221, App
576	27	56.2	720	4	US-09-394-272-14	Sequence 14, Appli	649	26	54.2	30	4	US-09-690-454-116	Sequence 116, App
577	27	56.2	734	4	US-09-726-219A-261	Sequence 261, App	650	26	54.2	51	4	US-09-583-110-3296	Sequence 3296, Ap
578	27	56.2	740	4	US-09-538-092-1178	Sequence 1178, Ap	651	26	54.2	66	4	US-09-248-796A-22537	Sequence 22537, A
579	27	56.2	809	4	US-09-543-681A-6686	Sequence 6686, Ap	652	26	54.2	66	4	US-09-248-796A-27118	Sequence 27118, A
580	27	56.2	838	4	US-09-252-991A-25809	Sequence 25809, A	653	26	54.2	67	4	US-09-513-999C-6385	Sequence 6385, Ap
581	27	56.2	850	4	US-09-902-540-12202	Sequence 12202, A	654	26	54.2	68	4	US-09-911-927-33	Sequence 33, Appli
582	27	56.2	864	4	US-09-134-000C-6025	Sequence 6025, Ap	655	26	54.2	68	4	US-09-911-888-33	Sequence 33, Appli
583	27	56.2	898	3	US-08-863-102-1	Sequence 1, Appli	656	26	54.2	69	3	US-09-025-151-9	Sequence 9, Appli
584	27	56.2	907	3	US-08-863-102-4	Sequence 4, Appli	657	26	54.2	69	3	US-08-965-762-33	Sequence 33, Appli
585	27	56.2	908	3	US-08-863-102-2	Sequence 2, Appli	658	26	54.2	69	4	US-09-637-240-9	Sequence 9, Appli
586	27	56.2	911	4	US-09-489-039A-13026	Sequence 13026, A	659	26	54.2	71	3	US-08-840-767-44	Sequence 44, Appli
587	27	56.2	917	4	US-08-259-451-11	Sequence 11, Appli	660	26	54.2	71	3	US-09-543-681A-4438	Sequence 4438, Ap
588	27	56.2	929	4	US-09-328-352-4591	Sequence 4591, Ap	661	26	54.2	72	4	US-09-543-681A-6642	Sequence 6642, Ap
589	27	56.2	950	3	US-09-449-285A-4	Sequence 4, Appli	662	26	54.2	75	4	US-08-905-223-277	Sequence 277, App
590	27	56.2	955	4	US-09-543-681A-6133	Sequence 6133, Ap	663	26	54.2	79	3	US-09-198-452A-1191	Sequence 1191, Ap
591	27	56.2	965	4	US-09-437-277-3	Sequence 3, Appli	664	26	54.2	83	4	US-09-621-976-4688	Sequence 4688, Ap
592	27	56.2	972	3	US-08-335-844A-24	Sequence 24, Appli	665	26	54.2	86	4	US-09-621-976-4688	Sequence 15889, A
593	27	56.2	973	3	US-09-129-366-24	Sequence 24, Appli	666	26	54.2	90	4	US-09-248-796A-15589	Sequence 11834, A
594	27	56.2	977	3	US-08-335-844A-22	Sequence 22, Appli	667	26	54.2	97	4	US-09-489-039A-11834	Sequence 6216, Ap
595	27	56.2	977	4	US-09-129-366-22	Sequence 22, Appli	668	26	54.2	98	4	US-09-513-999C-6216	Sequence 15680, A
596	27	56.2	987	4	US-09-949-016-10265	Sequence 10265, A	669	26	54.2	100	4	US-09-902-540-15680	Sequence 15680, A
597	27	56.2	1002	4	US-09-762-724-4	Sequence 4, Appli	670	26	54.2	101	3	US-09-518-046-15	Sequence 15, Appli
598	27	56.2	1032	3	US-09-115-954-8	Sequence 8, Appli	671	26	54.2	101	3	US-09-518-046-15	Sequence 15, Appli
599	27	56.2	1044	3	US-09-115-954-8	Sequence 8, Appli	672	26	54.2	101	3	US-09-248-796A-19931	Sequence 19931, A
600	27	56.2	1063	4	US-09-595-857B-29	Sequence 29, Appli	673	26	54.2	109	3	US-08-936-165A-440	Sequence 440, App
601	27	56.2	1081	4	US-09-529-239D-19	Sequence 19, Appli	674	26	54.2	109	4	US-09-673-395A-470	Sequence 470, App
602	27	56.2	1085	4	US-09-734-674-4	Sequence 4, Appli	675	26	54.2	109	4	US-09-673-395A-486	Sequence 486, App
603	27	56.2	1157	4	US-09-538-092-1328	Sequence 1328, Ap	676	26	54.2	109	4	US-09-673-395A-622	Sequence 622, App
604	27	56.2	1225	4	US-09-583-110-3637	Sequence 3637, Ap	677	26	54.2	109	4	US-09-513-999C-8128	Sequence 8128, Ap
605	27	56.2	1236	4	US-09-883-134-4	Sequence 4, Appli	678	26	54.2	112	4	US-09-893-737-66	Sequence 66, Appli
606	27	56.2	1239	4	US-09-107-433-4267	Sequence 4267, Ap	679	26	54.2	113	4	US-09-489-039A-11952	Sequence 11952, A
607	27	56.2	1260	4	US-09-328-352-6746	Sequence 6746, Ap	680	26	54.2	114	4	US-09-513-999C-5733	Sequence 5733, Ap
608	27	56.2	1475	4	US-09-538-092-1160	Sequence 1160, Ap	681	26	54.2	115	3	US-08-516-859A-100	Sequence 100, App
609	27	56.2	1618	3	US-08-462-467B-4	Sequence 4, Appli	682	26	54.2	115	3	US-09-586-472-100	Sequence 100, App
610	27	56.2	2008	4	US-09-091-501B-8	Sequence 8, Appli	683	26	54.2	115	4	US-09-528-706-100	Sequence 100, App
611	27	56.2	2325	3	US-08-417-089-6	Sequence 6, Appli	684	26	54.2	115	4	US-09-270-767-44981	Sequence 44981, A

685	26	54.2	118	4	US-09-543-681A-5800	Sequence 5800, Ap	758	26	54.2	252	4	US-09-543-681A-4245	Sequence 4245, Ap
686	26	54.2	119	1	US-07-828-853-1	Sequence 1, Appli	759	26	54.2	254	4	US-09-543-681A-7648	Sequence 7648, Ap
687	26	54.2	120	2	US-08-937-972-5	Sequence 5, Appli	760	26	54.2	255	4	US-09-134-001C-4116	Sequence 4116, Ap
688	26	54.2	126	4	US-08-248-796A-15590	Sequence 15590, A	761	26	54.2	256	4	US-09-902-540-13951	Sequence 13951, A
689	26	54.2	132	4	US-09-732-210-1313	Sequence 1313, Ap	762	26	54.2	260	4	US-09-134-001C-4872	Sequence 4872, Ap
690	26	54.2	132	4	US-09-732-210-1314	Sequence 1314, Ap	763	26	54.2	260	4	US-09-248-796A-17789	Sequence 17789, A
691	26	54.2	133	4	US-09-270-767-57154	Sequence 57154, A	764	26	54.2	260	4	US-09-248-796A-18584	Sequence 18584, A
692	26	54.2	134	4	US-09-621-976-40311	Sequence 40311, Ap	765	26	54.2	264	4	US-09-270-767-33115	Sequence 33115, A
693	26	54.2	138	4	US-09-902-540-10829	Sequence 10829, A	766	26	54.2	264	4	US-09-270-767-41787	Sequence 41787, A
694	26	54.2	139	4	US-09-328-352-7201	Sequence 7201, Ap	767	26	54.2	264	4	US-09-270-767-48332	Sequence 48332, A
695	26	54.2	142	4	US-09-328-352-5761	Sequence 5761, Ap	768	26	54.2	265	4	US-09-319-989-2	Sequence 2, Appli
696	26	54.2	143	4	US-09-134-000C-4059	Sequence 4059, A	769	26	54.2	265	4	US-09-328-352-4811	Sequence 4811, Ap
697	26	54.2	144	4	US-09-230-078A-6	Sequence 6, Appli	770	26	54.2	265	4	US-09-248-796A-16392	Sequence 16392, A
698	26	54.2	144	4	US-09-902-540-12977	Sequence 12977, A	771	26	54.2	265	4	US-09-538-092-325	Sequence 325, App
699	26	54.2	145	4	US-09-270-767-47454	Sequence 47454, A	772	26	54.2	266	4	US-09-134-000C-4815	Sequence 4815, Ap
700	26	54.2	146	4	US-09-270-767-32237	Sequence 32237, A	773	26	54.2	267	4	US-09-602-777A-156	Sequence 156, App
701	26	54.2	148	4	US-09-252-991A-18029	Sequence 18029, A	774	26	54.2	267	4	US-09-602-777A-158	Sequence 158, App
702	26	54.2	151	4	US-09-902-540-10111	Sequence 10111, A	775	26	54.2	269	4	US-09-134-001C-3461	Sequence 3461, Ap
703	26	54.2	155	4	US-09-270-767-32485	Sequence 32485, A	776	26	54.2	273	4	US-09-270-767-60531	Sequence 60531, A
704	26	54.2	155	4	US-09-270-767-47702	Sequence 47702, A	777	26	54.2	277	4	US-09-252-991A-29257	Sequence 29257, A
705	26	54.2	159	3	US-08-796-792-2	Sequence 2, Appli	778	26	54.2	277	4	US-09-198-452A-555	Sequence 555, App
706	26	54.2	159	4	US-09-198-452A-462	Sequence 462, App	779	26	54.2	278	4	US-09-438-185A-516	Sequence 516, App
707	26	54.2	159	4	US-09-491-795-2	Sequence 2, Appli	780	26	54.2	278	4	US-09-792-024-100	Sequence 100, App
708	26	54.2	160	4	US-09-252-991A-31843	Sequence 31843, A	781	26	54.2	278	4	US-09-902-540-16068	Sequence 16068, A
709	26	54.2	160	4	US-09-270-767-42694	Sequence 42694, A	782	26	54.2	279	4	US-09-134-000C-6430	Sequence 6430, Ap
710	26	54.2	165	1	US-08-233-788A-40	Sequence 40, Appl	783	26	54.2	279	4	US-09-270-767-35048	Sequence 35048, A
711	26	54.2	165	4	US-09-252-991A-27037	Sequence 27037, A	784	26	54.2	279	4	US-09-270-767-50265	Sequence 50265, A
712	26	54.2	166	4	US-09-248-796A-14783	Sequence 14783, A	785	26	54.2	279	4	US-09-710-279-858	Sequence 858, App
713	26	54.2	167	4	US-09-902-540-13832	Sequence 13832, A	786	26	54.2	279	4	US-09-925-637-50	Sequence 50, Appl
714	26	54.2	170	3	US-08-858-207A-519	Sequence 519, App	787	26	54.2	281	4	US-09-248-796A-17769	Sequence 17769, A
715	26	54.2	172	4	US-09-252-991A-26790	Sequence 26790, A	788	26	54.2	282	4	US-09-270-767-43326	Sequence 43326, A
716	26	54.2	173	4	US-09-902-540-11877	Sequence 11877, A	789	26	54.2	284	3	US-09-134-001C-3329	Sequence 3329, Ap
717	26	54.2	175	4	US-09-902-540-16269	Sequence 16269, A	790	26	54.2	285	4	US-09-489-039A-13043	Sequence 13043, A
718	26	54.2	176	1	US-08-449-922-1	Sequence 1, Appli	791	26	54.2	285	4	US-09-248-796A-15972	Sequence 15972, A
719	26	54.2	178	4	US-09-270-767-36123	Sequence 36123, A	792	26	54.2	287	4	US-09-252-991A-22466	Sequence 22466, A
720	26	54.2	178	4	US-09-270-767-51340	Sequence 51340, A	793	26	54.2	289	4	US-09-248-796A-19772	Sequence 19772, A
721	26	54.2	180	4	US-09-230-078A-4	Sequence 4, Appli	794	26	54.2	293	4	US-09-489-039A-7890	Sequence 7890, Ap
722	26	54.2	183	4	US-09-107-532A-4014	Sequence 4014, Ap	795	26	54.2	297	4	US-09-543-681A-8242	Sequence 8242, Ap
723	26	54.2	183	4	US-09-107-532A-5935	Sequence 5935, Ap	796	26	54.2	297	4	US-09-107-433-5074	Sequence 5074, Ap
724	26	54.2	186	4	US-09-248-796A-26092	Sequence 26092, A	797	26	54.2	298	4	US-09-270-767-45154	Sequence 45154, A
725	26	54.2	187	4	US-09-242-890-1	Sequence 1, Appli	798	26	54.2	299	4	US-09-949-016-8018	Sequence 8018, Ap
726	26	54.2	190	4	US-09-270-767-43995	Sequence 43995, A	799	26	54.2	302	4	US-09-270-767-59138	Sequence 59138, Ap
727	26	54.2	191	4	US-09-248-796A-20248	Sequence 20248, A	800	26	54.2	303	4	US-09-252-991A-27866	Sequence 27866, A
728	26	54.2	194	4	US-09-248-796A-14593	Sequence 14593, A	801	26	54.2	303	4	US-09-902-540-11322	Sequence 11322, A
729	26	54.2	199	2	US-08-831-575-8	Sequence 8, Appli	802	26	54.2	303	4	US-09-302-540-14454	Sequence 14454, A
730	26	54.2	199	2	US-08-831-575-9	Sequence 9, Appli	803	26	54.2	304	4	US-09-538-092-155	Sequence 155, App
731	26	54.2	199	4	US-09-710-279-1342	Sequence 1342, Ap	804	26	54.2	307	4	US-09-543-681A-6849	Sequence 6849, Ap
732	26	54.2	200	4	US-09-489-039A-11359	Sequence 11359, A	805	26	54.2	309	4	US-09-252-991A-18385	Sequence 18385, A
733	26	54.2	200	4	US-09-242-890-10	Sequence 10, Appl	806	26	54.2	309	4	US-09-603-208A-110	Sequence 110, App
734	26	54.2	203	4	US-09-252-991A-21364	Sequence 21364, A	807	26	54.2	310	1	US-08-433-783-42	Sequence 42, Appl
735	26	54.2	204	4	US-09-902-540-14744	Sequence 14744, A	808	26	54.2	310	2	US-08-337-358-42	Sequence 42, Appl
736	26	54.2	208	2	US-08-606-143-41	Sequence 41, Appl	809	26	54.2	310	4	US-09-270-767-47942	Sequence 47942, A
737	26	54.2	208	2	US-08-606-143-43	Sequence 43, Appl	810	26	54.2	310	5	PCT-US95-07537A-42	Sequence 42, Appl
738	26	54.2	213	4	US-09-107-532A-3797	Sequence 3797, Ap	811	26	54.2	310	5	PCT-US95-07537-42	Sequence 42, Appl
739	26	54.2	213	4	US-09-270-767-56727	Sequence 56727, A	812	26	54.2	311	4	US-09-583-110-2874	Sequence 2874, Ap
740	26	54.2	217	4	US-09-650-454-112	Sequence 112, App	813	26	54.2	313	3	US-09-336-643A-81	Sequence 81, Appl
741	26	54.2	219	4	US-09-484-577A-32	Sequence 32, Appl	814	26	54.2	313	4	US-09-561-763-8	Sequence 8, Appli
742	26	54.2	220	4	US-09-252-991A-16651	Sequence 16651, A	815	26	54.2	313	4	US-09-431-367B-8	Sequence 8, Appli
743	26	54.2	224	4	US-09-229-583A-2	Sequence 2, Appli	816	26	54.2	315	4	US-09-248-796A-19532	Sequence 19532, A
744	26	54.2	224	4	US-09-242-890-11	Sequence 11, Appl	817	26	54.2	317	2	US-08-864-799-5	Sequence 5, Appli
745	26	54.2	224	4	US-10-187-904-2	Sequence 2, Appli	818	26	54.2	317	4	US-09-949-016-11333	Sequence 11333, A
746	26	54.2	224	4	US-09-248-796A-14183	Sequence 14183, A	819	26	54.2	318	1	US-08-599-480-2	Sequence 2, Appli
747	26	54.2	224	4	US-09-248-796A-15240	Sequence 15240, A	820	26	54.2	318	2	US-08-842-199-2	Sequence 2, Appli
748	26	54.2	226	4	US-09-248-796A-20804	Sequence 20804, A	821	26	54.2	318	4	US-09-172-070-2	Sequence 2, Appli
749	26	54.2	235	4	US-09-252-991A-29814	Sequence 29814, A	822	26	54.2	318	5	PCT-US95-00190-2	Sequence 2, Appli
750	26	54.2	243	4	US-09-489-039A-13870	Sequence 13870, A	823	26	54.2	320	4	US-09-526-400-2	Sequence 2, Appli
751	26	54.2	244	4	US-09-710-279-566	Sequence 566, App	824	26	54.2	322	4	US-09-724-797-56	Sequence 56, Appli
752	26	54.2	246	3	US-09-134-001C-5292	Sequence 5292, Ap	825	26	54.2	323	4	US-09-438-185A-443	Sequence 443, App
753	26	54.2	247	4	US-09-248-796A-14828	Sequence 14828, Ap	826	26	54.2	325	4	US-09-134-000C-4346	Sequence 4346, Ap
754	26	54.2	249	4	US-09-270-767-38338	Sequence 38338, A	827	26	54.2	328	4	US-09-270-767-59817	Sequence 59817, A
755	26	54.2	249	4	US-09-270-767-53555	Sequence 53555, A	828	26	54.2	335	4	US-09-252-991A-22013	Sequence 22013, A
756	26	54.2	251	4	US-09-710-279-3050	Sequence 3050, Ap	829	26	54.2	336	4	US-09-248-796A-17499	Sequence 17499, A
757	26	54.2	251	4	US-10-000-489-26	Sequence 26, Appl	830	26	54.2	336	4	US-09-949-016-7218	Sequence 7218, Ap

831	26	54.2	337	4	US-09-583-110-3391	Sequence 3391, Ap	904	26	54.2	433	4	US-09-949-016-7709	Sequence 7709, Ap
832	26	54.2	338	1	US-08-891-254-1	Sequence 1, Appli	905	26	54.2	434	4	US-09-949-016-8735	Sequence 8735, Ap
833	26	54.2	338	2	US-08-484-358-2	Sequence 2, Appli	906	26	54.2	436	4	US-09-252-991A-18821	Sequence 18821, A
834	26	54.2	338	2	US-08-819-539-1	Sequence 1, Appli	907	26	54.2	445	3	US-08-974-691-6	Sequence 6, Appli
835	26	54.2	338	2	US-09-030-270A-1	Sequence 1, Appli	908	26	54.2	451	1	US-08-154-365-2	Sequence 2, Appli
836	26	54.2	338	3	US-09-118-959-2	Sequence 2, Appli	909	26	54.2	451	1	US-08-453-117-2	Sequence 2, Appli
837	26	54.2	338	3	US-08-984-207-1	Sequence 1, Appli	910	26	54.2	451	2	US-08-948-222-2	Sequence 2, Appli
838	26	54.2	338	3	US-09-013-587-1	Sequence 1, Appli	911	26	54.2	451	2	US-08-973-145-2	Sequence 10, Appli
839	26	54.2	338	3	US-09-086-118-21	Sequence 21, Appli	912	26	54.2	451	3	US-09-276-400-10	Sequence 2, Appli
840	26	54.2	338	4	US-09-431-614-1	Sequence 1, Appli	913	26	54.2	451	3	US-08-974-691-2	Sequence 10, Appli
841	26	54.2	338	5	PCT-US96-08819-1	Sequence 1, Appli	914	26	54.2	451	3	US-09-448-076-10	Sequence 10, Appli
842	26	54.2	340	1	US-07-828-700-8	Sequence 8, Appli	915	26	54.2	451	3	US-09-702-572-10	Sequence 10, Appli
843	26	54.2	340	3	US-09-360-779-2	Sequence 2, Appli	916	26	54.2	451	3	US-09-270-767-45139	Sequence 45139, A
844	26	54.2	340	3	US-09-435-335-2	Sequence 2, Appli	917	26	54.2	451	5	PCT-US96-08081-2	Sequence 2, Appli
845	26	54.2	343	4	US-09-328-352-6577	Sequence 6577, Ap	918	26	54.2	453	6	5510466-4	Patent No. 5510466
846	26	54.2	344	4	US-09-724-623-79	Sequence 79, Appli	919	26	54.2	453	6	5510466-4	Patent No. 5510466
847	26	54.2	345	4	US-09-214-631-5	Sequence 5, Appli	920	26	54.2	454	3	US-09-240-639-11	Sequence 11, Appli
848	26	54.2	346	1	US-08-213-403-2	Sequence 2, Appli	921	26	54.2	454	4	US-09-538-092-1156	Sequence 11, Appli
849	26	54.2	346	1	US-08-458-077-2	Sequence 2, Appli	922	26	54.2	454	4	US-09-908-510A-11	Sequence 11, Appli
850	26	54.2	346	1	US-08-460-741-2	Sequence 2, Appli	923	26	54.2	454	4	US-09-905-744B-11	Sequence 11, Appli
851	26	54.2	346	1	US-08-747-240-2	Sequence 2, Appli	924	26	54.2	454	4	US-10-107-660-11	Sequence 11, Appli
852	26	54.2	346	1	US-08-299-567-6	Sequence 6, Appli	925	26	54.2	454	4	US-10-107-576-11	Sequence 11, Appli
853	26	54.2	346	4	US-09-039-642B-2	Sequence 2, Appli	926	26	54.2	454	4	US-09-905-732B-11	Sequence 11, Appli
854	26	54.2	346	4	US-08-635-130A-9	Sequence 9, Appli	927	26	54.2	454	4	US-09-905-743B-11	Sequence 11, Appli
855	26	54.2	347	4	US-09-949-016-9849	Sequence 9849, Ap	928	26	54.2	456	4	US-09-134-000C-6463	Sequence 6463, Ap
856	26	54.2	347	4	US-08-462-467B-18	Sequence 18, Appli	929	26	54.2	456	4	US-09-248-796A-17048	Sequence 17048, A
857	26	54.2	352	1	US-08-482-577B-18	Sequence 18, Appli	930	26	54.2	462	4	US-09-976-594-427	Sequence 427, App
858	26	54.2	352	3	US-08-289-222E-4	Sequence 4, Appli	931	26	54.2	469	1	US-08-313-288B-15	Sequence 15, Appli
859	26	54.2	352	3	US-09-218-176-2	Sequence 2, Appli	932	26	54.2	470	4	US-09-886-319A-35	Sequence 35, Appli
860	26	54.2	352	3	US-09-054-526B-4	Sequence 4, Appli	933	26	54.2	473	3	US-08-914-375C-71	Sequence 71, Appli
861	26	54.2	352	4	US-09-981-490B-3	Sequence 3, Appli	934	26	54.2	476	4	US-09-107-532A-5076	Sequence 5076, Ap
862	26	54.2	354	3	US-09-134-001C-5588	Sequence 5588, Ap	935	26	54.2	477	1	US-07-791-936A-2	Sequence 2, Appli
863	26	54.2	354	4	US-09-107-433-3389	Sequence 3389, Ap	936	26	54.2	477	1	US-08-383-781B-2	Sequence 2, Appli
864	26	54.2	356	4	US-09-949-016-10384	Sequence 10384, A	937	26	54.2	477	4	US-07-969-267B-2	Sequence 2, Appli
865	26	54.2	363	4	US-09-248-796A-15286	Sequence 15286, A	938	26	54.2	477	4	US-09-168-510-2	Sequence 2, Appli
866	26	54.2	365	4	US-09-134-000C-4369	Sequence 4369, Ap	939	26	54.2	477	4	US-09-826-509-493	Sequence 493, App
867	26	54.2	368	4	US-09-252-991A-33020	Sequence 33020, A	940	26	54.2	478	4	US-09-107-532A-6090	Sequence 6090, Ap
868	26	54.2	369	4	US-09-252-991A-20245	Sequence 20245, A	941	26	54.2	479	4	US-09-543-681A-7331	Sequence 7331, A
869	26	54.2	369	4	US-10-138-701-36	Sequence 36, Appli	942	26	54.2	480	4	US-09-252-991A-24951	Sequence 24951, A
870	26	54.2	369	4	US-09-902-540-12050	Sequence 12050, A	943	26	54.2	482	4	US-09-538-092-300	Sequence 200, App
871	26	54.2	370	4	US-09-252-991A-23998	Sequence 23998, A	944	26	54.2	484	4	US-09-949-016-9698	Sequence 9698, Ap
872	26	54.2	371	3	US-09-043-627-10	Sequence 10, Appli	945	26	54.2	490	1	US-08-294-770A-2	Sequence 2, Appli
873	26	54.2	376	4	US-09-270-767-42614	Sequence 42614, A	946	26	54.2	491	2	US-08-448-735C-2	Sequence 2, Appli
874	26	54.2	386	4	US-09-489-039A-8324	Sequence 8324, Ap	947	26	54.2	491	4	US-09-902-540-13995	Sequence 13995, A
875	26	54.2	391	4	US-09-270-767-44694	Sequence 44694, A	948	26	54.2	493	4	US-09-902-540-11934	Sequence 11934, A
876	26	54.2	392	4	US-09-489-039A-9827	Sequence 9827, Ap	949	26	54.2	497	4	US-09-134-000C-4549	Sequence 4549, Ap
877	26	54.2	392	4	US-09-902-540-12860	Sequence 12860, A	950	26	54.2	499	4	US-09-328-352-5415	Sequence 5415, Ap
878	26	54.2	395	1	US-08-723-938-3	Sequence 3, Appli	951	26	54.2	501	4	US-09-270-767-43743	Sequence 43743, A
879	26	54.2	395	2	US-09-080-538-3	Sequence 3, Appli	952	26	54.2	509	4	US-09-252-991A-23445	Sequence 23445, A
880	26	54.2	395	4	US-09-387-413-3	Sequence 3, Appli	953	26	54.2	518	3	US-09-065-383-27	Sequence 27, Appli
881	26	54.2	398	4	US-09-252-991A-21686	Sequence 21686, A	954	26	54.2	522	4	US-09-949-016-11189	Sequence 11189, A
882	26	54.2	400	4	US-09-691-270A-14	Sequence 14, Appli	955	26	54.2	542	1	US-08-246-583-3	Sequence 3, Appli
883	26	54.2	402	1	US-08-403-545-2	Sequence 2, Appli	956	26	54.2	542	4	US-09-636-791A-5	Sequence 5, Appli
884	26	54.2	409	3	US-08-404-381-2	Sequence 2, Appli	957	26	54.2	542	4	US-09-538-092-672	Sequence 8, Appli
885	26	54.2	412	4	US-09-538-092-709	Sequence 709, App	958	26	54.2	547	1	US-08-083-948-8	Sequence 8, Appli
886	26	54.2	414	4	US-09-543-681A-6647	Sequence 6647, Ap	959	26	54.2	547	1	US-08-475-694-8	Sequence 8, Appli
887	26	54.2	415	4	US-09-538-092-708	Sequence 708, App	960	26	54.2	547	1	US-08-713-057-8	Sequence 8, Appli
888	26	54.2	416	4	US-09-270-767-45205	Sequence 45205, A	961	26	54.2	547	3	US-09-347-078-36	Sequence 36, Appli
889	26	54.2	417	3	US-08-462-467B-6	Sequence 6, Appli	962	26	54.2	553	3	US-08-630-172-13	Sequence 13, Appli
890	26	54.2	417	3	US-08-462-467B-10	Sequence 10, Appli	963	26	54.2	553	3	US-09-375-419-13	Sequence 13, Appli
891	26	54.2	417	3	US-08-462-467B-16	Sequence 16, Appli	964	26	54.2	553	4	US-09-465-559-14	Sequence 14, Appli
892	26	54.2	417	4	US-09-248-796A-16038	Sequence 16038, A	965	26	54.2	553	4	US-09-902-540-15181	Sequence 15181, A
893	26	54.2	419	3	US-09-064-411A-7	Sequence 7, Appli	966	26	54.2	554	3	US-08-462-467B-22	Sequence 22, Appli
894	26	54.2	420	3	US-09-008-271A-4	Sequence 4, Appli	967	26	54.2	554	4	US-09-538-092-245	Sequence 245, App
895	26	54.2	420	3	US-08-974-691-8	Sequence 8, Appli	968	26	54.2	556	3	US-08-462-467B-24	Sequence 24, Appli
896	26	54.2	420	4	US-09-705-448-1	Sequence 1, Appli	969	26	54.2	556	3	US-08-462-467B-26	Sequence 26, Appli
897	26	54.2	421	4	US-09-902-540-10569	Sequence 10569, A	970	26	54.2	556	3	US-09-902-540-12461	Sequence 12461, A
898	26	54.2	421	4	US-09-248-796A-17203	Sequence 17203, A	971	26	54.2	556	4	US-09-252-991A-17202	Sequence 17202, A
899	26	54.2	428	3	US-09-134-001C-4879	Sequence 4879, Ap	972	26	54.2	558	4	US-09-489-039A-9311	Sequence 9311, Ap
900	26	54.2	430	4	US-09-248-796A-16772	Sequence 16772, A	973	26	54.2	561	2	US-08-532-795-2	Sequence 2, Appli
901	26	54.2	433	3	US-09-697-367-22	Sequence 22, Appli	974	26	54.2	562	2	US-08-532-795-2	Sequence 2, Appli
902	26	54.2	433	4	US-09-705-448-3	Sequence 3, Appli	975	26	54.2	566	2	US-08-484-993B-41	Sequence 41, Appli
903	26	54.2	433	4	US-09-918-909A-22	Sequence 22, Appli	976	26	54.2	566	2	US-08-484-158B-41	Sequence 41, Appli

977 26 54.2 566 2 US-08-484-596A-41 Sequence 41, Appl  
978 26 54.2 566 2 US-08-480-150A-41 Sequence 41, Appl  
979 26 54.2 566 3 US-08-458-731-41 Sequence 41, Appl  
980 26 54.2 566 3 US-08-149-223A-41 Sequence 41, Appl  
981 26 54.2 566 4 US-09-248-796A-20411 Sequence 20411, A  
982 26 54.2 569 2 US-08-532-795-23 Sequence 23, Appl  
983 26 54.2 569 2 US-08-532-795-29 Sequence 29, Appl  
984 26 54.2 570 2 US-08-532-795-25 Sequence 25, Appl  
985 26 54.2 571 2 US-08-532-795-19 Sequence 19, Appl  
986 26 54.2 571 2 US-08-532-795-21 Sequence 21, Appl  
987 26 54.2 574 2 US-08-532-795-27 Sequence 27, Appl  
988 26 54.2 577 4 US-09-902-540-15452 Sequence 15452, A  
989 26 54.2 577 4 US-09-902-540-15872 Sequence 15872, A  
990 26 54.2 583 2 US-08-616-392C-4 Sequence 4, Appl  
991 26 54.2 592 4 US-09-614-912-90 Sequence 90, Appl  
992 26 54.2 605 4 US-09-252-991A-18837 Sequence 901, Ap  
993 26 54.2 613 4 US-09-949-016-9601 Sequence 4, Appl  
994 26 54.2 614 1 US-08-262-338A-4 Sequence 4, Appl  
995 26 54.2 614 1 US-08-460-114A-4 Sequence 4, Appl  
996 26 54.2 614 3 US-09-347-878-38 Sequence 38, Appl  
997 26 54.2 614 4 US-09-133-966A-4 GENERAL INFORMA  
998 26 54.2 618 1 US-08-468-853-10 Sequence 10, Appl  
999 26 54.2 618 1 US-08-468-855-10 Sequence 10, Appl  
1000 26 54.2 618 1 US-08-310-357-10 Sequence 10, Appl

## ALIGNMENTS

RESULT 1  
US-09-346-860-9  
; Sequence 9, Application US/09346860  
; Patent No. 6221641  
; GENERAL INFORMATION:  
; APPLICANT: Khosla, Chaitan  
; APPLICANT: Lau, Janice  
; APPLICANT: Pohl, Nicola L.  
; TITLE OF INVENTION: METHOD FOR MAKING POLYKETIDES  
; FILE REFERENCE: 30062-20025.00  
; CURRENT APPLICATION NUMBER: US/09/346,860  
; CURRENT FILING DATE: 1999-07-02  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 50  
; TYPE: PRT  
; ORGANISM: Streptomyces sp.  
US-09-346-860-9

Query Match 75.0%; Score 36; DB 3; Length 50;  
Best Local Similarity 66.7%; Pred. No. 2.6;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
Db 5 GPHETLTL 13

RESULT 2  
US-09-735-685-9  
; Sequence 9, Application US/09735685  
; Patent No. 6551802  
; GENERAL INFORMATION:  
; APPLICANT: Khosla, Chaitan  
; APPLICANT: Lau, Janice  
; APPLICANT: Pohl, Nicola L.  
; TITLE OF INVENTION: METHOD FOR MAKING POLYKETIDES  
; FILE REFERENCE: 30062-20025.00  
; CURRENT APPLICATION NUMBER: US/09/735,685  
; CURRENT FILING DATE: 2000-12-12  
; PRIOR APPLICATION NUMBER: US/09/346,860  
; PRIOR FILING DATE: 1999-07-02  
; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 50  
; TYPE: PRT  
; ORGANISM: Streptomyces sp.  
US-09-735-685-9

Query Match 75.0%; Score 36; DB 4; Length 50;  
Best Local Similarity 66.7%; Pred. No. 2.6;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
Db 5 GPHETLTL 13

RESULT 3  
US-08-905-223-493  
; Sequence 493, Application US/08905223  
; Patent No. 6222029  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Jean-Baptiste D.  
; APPLICANT: Duelet, Aymeric  
; APPLICANT: Lacroix, Bruno  
; TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS  
; NUMBER OF SEQUENCES: 503  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 501 West Broadway  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92101-3505  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy Disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: Win95  
; SOFTWARE: Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/905,223  
; FILING DATE:  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelsen, Ned A.  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 493:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 79 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
; ORIGINAL SOURCE:  
; ORGANISM: Homo Sapiens  
; TISSUE TYPE: Brain  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: -18...-1  
; IDENTIFICATION METHOD: Von Heijne matrix  
; OTHER INFORMATION: score 3.8  
; OTHER INFORMATION: seq GXALGLPSLAKA/ED  
US-08-905-223-493

Query Match 75.0%; Score 36; DB 3; Length 79;  
Best Local Similarity 66.7%; Pred. No. 4.5;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
Db 54 GPHETLXL 62

```

RESULT 4
US-10-127-641-5
; Sequence 5, Application US/10127641
; Patent No. 6824778
; GENERAL INFORMATION:
; APPLICANT: Hart, Mary Kate
; TITLE OF INVENTION: Prophylactic and Therapeutic Monoclonal Antibodies
; FILE REFERENCE: 003/251/SAP
; CURRENT APPLICATION NUMBER: US/10/127,641
; PRIOR FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: US 60/285,601
; PRIOR FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 5
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Eastern equine encephalitis virus
US-10-127-641-5
Query Match 72.9%; Score 35; DB 4; Length 58;
Best Local Similarity 75.0%; Pred. No. 5;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PHETITAL 9
Db 30 PHETLTM 37

RESULT 5
US-09-248-796A-19847
; Sequence 19847, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19847
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19847
Query Match 72.9%; Score 35; DB 4; Length 91;
Best Local Similarity 87.5%; Pred. No. 8.4;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PHETITAL 9
Db 62 PHEVITAL 69

RESULT 6
US-08-997-080-154
; Sequence 154, Application US/08997080
; Patent No. 5968524
; GENERAL INFORMATION:
; APPLICANT: WATSON, JAMES D.
; APPLICANT: TAN, PAUL L.J.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman

```

```

; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,080
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1007
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 154:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 748 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-997-080-154
Query Match 72.9%; Score 35; DB 2; Length 748;
Best Local Similarity 77.8%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPHETITAL 9
Db 405 GPHETRAL 413

RESULT 7
US-08-997-362-154
; Sequence 154, Application US/08997362
; Patent No. 5985287
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Hiyama, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Scott, Linda
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,362
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

```
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
; FILING DATE: June 12, 1997
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
; FILING DATE: August 29, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 154:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 748 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-997-362-154

Query Match 72.9%; Score 35; DB 2; Length 748;
Best Local Similarity 77.8%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9
Db 405 GPHETHRAL 413

RESULT 8
US-09-095-855-154
; Sequence 154, Application US/09095855
; Patent No. 6160093
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Compounds and Methods for
; TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/095,855
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/705,347
; FILING DATE: 29-AUG-1996
; APPLICATION NUMBER: 08/873,970
; FILING DATE: 12-JUN-1997
; APPLICATION NUMBER: 08/997,362
; FILING DATE: 23-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 154:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 748 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-095-855-154

Query Match 72.9%; Score 35; DB 3; Length 748;
Best Local Similarity 77.8%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9
Db 405 GPHETHRAL 413

RESULT 9
US-09-324-542-154
; Sequence 154, Application US/09324542
; Patent No. 6328978
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L.J.
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
; FILE REFERENCE: 11000.1007c1
; CURRENT APPLICATION NUMBER: US/09/324,542
; CURRENT FILING DATE: 1999-06-02
; EARLIER APPLICATION NUMBER: US 08/997,080
; EARLIER FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 154
; LENGTH: 748
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (119)...(119)
; US-09-324-542-154

Query Match 72.9%; Score 35; DB 3; Length 748;
Best Local Similarity 77.8%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9
Db 405 GPHETHRAL 413

RESULT 10
US-09-205-426-154
; Sequence 154, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Compounds and Methods for Treatment and
; TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
; FILE REFERENCE: 11000.1002c4
; CURRENT APPLICATION NUMBER: US/09/205,426
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: 09/095,855
; EARLIER FILING DATE: 1998-06-11
; EARLIER APPLICATION NUMBER: 08/997,362
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/873,970
; EARLIER FILING DATE: 1997-06-12
; EARLIER APPLICATION NUMBER: 08/705,347
; EARLIER FILING DATE: 1996-08-29
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
```



; SEQ ID NO 154  
; LENGTH: 748  
; TYPE: PRT  
; ORGANISM: Mycobacterium vaccae  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (119)...(119)  
US-09-205-426-154

Query Match 72.9%; Score 35; DB 4; Length 748;  
Best Local Similarity 77.8%; Pred. No. 94;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPHETITAL 9  
||| ||| ||  
Db 405 GPHETITAL 413

RESULT 11  
US-09-005-180A-3  
; Sequence 3, Application US/09005180A  
; Patent No. 6124446  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: HUMAN VPS35/MEM3-RELATED PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/005,180A  
; FILING DATE: Filed January 8, 1998  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PP-0457 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 754 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 1354049  
US-09-005-180A-3

Query Match 72.9%; Score 35; DB 3; Length 754;  
Best Local Similarity 75.0%; Pred. No. 95;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PHETITAL 9  
||| ||| ||  
Db 523 PHETITAL 530

RESULT 12  
US-09-513-999C-6366  
; Sequence 6366, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; Patent No. 6783961  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 6366  
; LENGTH: 102  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: 13  
; OTHER INFORMATION: Xaa=Ala or Glu  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: 22  
; OTHER INFORMATION: Xaa=Ala or Pro or Ser or Thr  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: 88  
; OTHER INFORMATION: Xaa=Cys or Ser  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: 98  
; OTHER INFORMATION: Xaa=Gly or Val  
US-09-513-999C-6366

Query Match 70.8%; Score 34; DB 4; Length 102;  
Best Local Similarity 75.0%; Pred. No. 15;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPHETITA 8  
||| ||| ||  
Db 4 GPHSTATA 11

RESULT 13  
US-09-615-192A-311  
; Sequence 311, Application US/09615192A  
; Patent No. 6410718  
; GENERAL INFORMATION:  
; APPLICANT: Bloksberg, Leonard N.  
; APPLICANT: Havukkala, Ilkka  
; TITLE OF INVENTION: Materials and Methods for the  
; TITLE OF INVENTION: Modification of Plant Lignin Content  
; FILE REFERENCE: 11000.1003c4U  
; CURRENT APPLICATION NUMBER: US/09/615,192A  
; CURRENT FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 08/975,316  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: US 08/713,000  
; PRIOR FILING DATE: 1996-09-11  
; PRIOR APPLICATION NUMBER: US 09/169,789  
; PRIOR FILING DATE: 1998-10-09  
; NUMBER OF SEQ ID NOS: 405  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 311  
; LENGTH: 154  
; TYPE: PRT

; ORGANISM: Pinus radiata  
US-09-615-192A-311

Query Match 70.8%; Score 34; DB 4; Length 154;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PHETITAL 9  
||:||||  
Db 61 PHQAITAL 68

## RESULT 14

US-09-134-000C-5791  
; Sequence 5791, Application US/09134000C  
; Patent No. 6617156

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 5791  
; LENGTH: 300  
; TYPE: PRT

; ORGANISM: Enterococcus faecalis  
US-09-134-000C-5791

Query Match 70.8%; Score 34; DB 4; Length 300;  
Best Local Similarity 62.5%; Pred. No. 53;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 PHETITAL 9  
||:||||  
Db 28 PHQTLTSL 35

## RESULT 15

US-09-252-991A-17077  
; Sequence 17077, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17077  
; LENGTH: 699  
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17077

Query Match 70.8%; Score 34; DB 4; Length 699;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
|||:||||  
Db 347 GPHQPIVAL 355

## RESULT 16

US-09-543-681A-5706  
; Sequence 5706, Application US/09543681A  
; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 5706  
; LENGTH: 717  
; TYPE: PRT

; ORGANISM: Proteus mirabilis  
US-09-543-681A-5706

Query Match 70.8%; Score 34; DB 4; Length 717;  
Best Local Similarity 77.8%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
|||:||||  
Db 497 GVHETINAL 505

## RESULT 17

US-09-107-532A-4275  
; Sequence 4275, Application US/09107532A  
; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4275:

SEQUENCE CHARACTERISTICS:  
LENGTH: 762 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

```

; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...762
; SEQUENCE DESCRIPTION: SEQ ID NO: 4275:
US-09-107-532A-4275
Query Match 70.8%; Score 34; DB 4; Length 762;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPHETITAL 9
Db 383 GPHELINSL 391

RESULT 18
US-08-936-135-4
; Sequence 4, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2584 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-936-135-4
Query Match 70.8%; Score 34; DB 3; Length 2584;
Best Local Similarity 66.7%; Pred. No. 6.2e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPHETITAL 9
Db 1165 GPHGLVAL 1173

RESULT 19
US-08-936-135-2
; Sequence 2, Application US/08936135
; Patent No. 6054293
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: He, Zhigang
; APPLICANT: Chen, Hang
; TITLE OF INVENTION: Semaphorin Receptors
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/936,135
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC97-288-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2584 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-936-135-2
Query Match 70.8%; Score 34; DB 3; Length 2588;
Best Local Similarity 66.7%; Pred. No. 6.2e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPHETITAL 9
Db 1167 GPHGLVAL 1175

RESULT 20
US-09-107-532A-7025
; Sequence 7025, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION NUMBER: 60/085,598
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
```

```
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 7025:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 81 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...81
; SEQUENCE DESCRIPTION: SEQ ID NO: 7025:
US-09-107-532A-7025

Query Match 68.8%; Score 33; DB 4; Length 81;
Best Local Similarity 75.0%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PHETITAL 9
Db 37 PHMTITAI 44

RESULT 21
US-09-252-991A-22552
; Sequence 22552, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22552
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22552

Query Match 68.8%; Score 33; DB 4; Length 364;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPHETITA 8
Db 31 GPHRTSTA 38

RESULT 22
US-09-248-796A-19514
; Sequence 19514, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
```

```
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19514
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (87)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-248-796A-19514

Query Match 68.8%; Score 33; DB 4; Length 401;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PHETITAL 9
Db 114 PHSTISAL 121

RESULT 23
US-07-832-855-2
; Sequence 2, Application US/07832855
; Patent No. 5478727
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; APPLICANT: Liu, Fenyoung
; TITLE OF INVENTION: Methods and Compositions of a
; TITLE OF INVENTION: Preparation and Use of A Herpes Protease
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: 321 No. 5478727th Clark Street, Suite 800
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/832,855
; FILING DATE: 19920207
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coolley, Ronald B.
; REGISTRATION NUMBER: 27,187
; REFERENCE/DOCKET NUMBER: ARCD045
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 744-0090
; TELEFAX: (312) 245-4961
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-832-855-2

Query Match 68.8%; Score 33; DB 1; Length 635;
Best Local Similarity 77.8%; Pred. No. 2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPHETITAL 9
Db 473 GTNETITAL 481
```

RESULT 24  
US-08-176-320-2  
; Sequence 2, Application US/08176320  
; Patent No. 6410704  
; GENERAL INFORMATION:  
; APPLICANT: Roizman, Bernard  
; APPLICANT: Liu, Fenyong  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
; TITLE OF INVENTION: PREPARATION AND USE OF A HERPES PROTEASE  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Alice O. Martin  
; STREET: 321 No. 6410704th Clark Street, Suite 800  
; CITY: Chicago  
; STATE: ILL  
; COUNTRY: USA  
; ZIP: 60610  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/176.320  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/705.814  
; FILING DATE: 24-MAY-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cooley, Ronald B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312)744-0090  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 635 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-176-320-2

Query Match 68.8%; Score 33; DB 4; Length 635;  
Best Local Similarity 77.8%; Pred. No. 2e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GPHETITAL 9  
|:|||||  
Db 473 GNETITAL 481

RESULT 25  
US-09-489-039A-8750  
; Sequence 8750, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 8750  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-8750

Query Match 68.8%; Score 33; DB 4; Length 726;

Best Local Similarity 77.8%; Pred. No. 2.3e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPHETITAL 9  
|:|||||  
Db 507 GINETINAL 515

RESULT 26  
US-09-762-724-10  
; Sequence 10, Application US/09762724  
; Patent No. 6664053  
; GENERAL INFORMATION:  
; APPLICANT: Kovacs, et al.  
; TITLE OF INVENTION: Identification of a region of the major surface  
; TITLE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii  
; FILE REFERENCE: 4239-58054  
; CURRENT APPLICATION NUMBER: US/09/762,724  
; CURRENT FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: PCT/US99/18750  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: US 60/096,805  
; PRIOR FILING DATE: 1998-08-17  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 1009  
; TYPE: PRT  
; ORGANISM: Pneumocystis carinii sp. f. hominis  
US-09-762-724-10

Query Match 68.8%; Score 33; DB 4; Length 1009;  
Best Local Similarity 83.3%; Pred. No. 3.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PHETIT 7  
|:|||||  
Db 921 PHETVT 926

RESULT 27  
US-09-762-724-12  
; Sequence 12, Application US/09762724  
; Patent No. 6664053  
; GENERAL INFORMATION:  
; APPLICANT: Kovacs, et al.  
; TITLE OF INVENTION: Identification of a region of the major surface  
; TITLE OF INVENTION: glycoprotein (MSG) gene of human Pneumocystis carinii  
; FILE REFERENCE: 4239-58054  
; CURRENT APPLICATION NUMBER: US/09/762,724  
; CURRENT FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: PCT/US99/18750  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: US 60/096,805  
; PRIOR FILING DATE: 1998-08-17  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 1017  
; TYPE: PRT  
; ORGANISM: Pneumocystis carinii sp. f. hominis  
US-09-762-724-12

Query Match 68.8%; Score 33; DB 4; Length 1017;  
Best Local Similarity 83.3%; Pred. No. 3.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PHETIT 7  
|:|||||  
Db 912 PHETVT 917

RESULT 28

```
US-09-762-724-14
; Sequence 14, Application US/09762724
; Patent No. 6664053
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/09/762,724
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Pneumocystis carinii sp. f. hominis
US-09-762-724-14
Query Match 68.8%; Score 33; DB 4; Length 1023;
Best Local Similarity 83.3%; Pred. No. 3.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PHETIT 7
Db 916 PHETVT 921

RESULT 29
US-09-762-724-8
; Sequence 8, Application US/09762724
; Patent No. 6664053
; GENERAL INFORMATION:
; APPLICANT: Kovacs, et al.
; TITLE OF INVENTION: Identification of a region of the major surface
; FILE REFERENCE: 4239-58054
; CURRENT APPLICATION NUMBER: US/09/762,724
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: PCT/US99/18750
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: US 60/096,805
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1027
; TYPE: PRT
; ORGANISM: Pneumocystis carinii sp. f. hominis
US-09-762-724-8
Query Match 68.8%; Score 33; DB 4; Length 1027;
Best Local Similarity 83.3%; Pred. No. 3.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PHETIT 7
Db 922 PHETVT 927

RESULT 30
US-09-513-999C-5317
; Sequence 5317, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
```

```
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 5317
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 46
; OTHER INFORMATION: Xaa=Glu or Gly
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 48
; OTHER INFORMATION: Xaa=Lys or Asn
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 55
; OTHER INFORMATION: Xaa=Asp or Glu
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 88
; OTHER INFORMATION: Xaa=Ile or Met
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 91
; OTHER INFORMATION: Xaa=Ala or Pro
US-09-513-999C-5317
Query Match 66.7%; Score 32; DB 4; Length 108;
Best Local Similarity 85.7%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPHETIT 7
Db 96 GPHETIT 102

RESULT 31
US-09-007-484-4
; Sequence 4, Application US/09007484
; Patent No. 6072032
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: No. 6072032el Ftsy
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/007,484
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10081
```

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-994-2488  
TELEFAX: 215-994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 277 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-007-484-4

Query Match 66.7%; Score 32; DB 3; Length 277;  
Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PHETITAL 9  
|||||  
Db 192 PHETFLAL 199

## RESULT 32

US-09-309-682-4

Sequence 4, Application US/09309682  
Patent No. 6214348

## GENERAL INFORMATION:

APPLICANT: Black, Michael T.  
TITLE OF INVENTION: No. 6214348el Ftsy

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads

STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

CITY: Philadelphia

STATE: PA

COUNTRY: US

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/309,682

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/007,484

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Falk, Stephen T

REGISTRATION NUMBER: 36,795

REFERENCE/DOCKET NUMBER: GM10081

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-994-2488

TELEFAX: 215-994-2222

TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 277 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-309-682-4

Query Match 66.7%; Score 32; DB 3; Length 277;  
Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PHETITAL 9  
|||||  
Db 192 PHETFLAL 199

## RESULT 33

US-09-335-409-22

Sequence 22, Application US/09335409

Patent No. 6121029

## GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James

APPLICANT: Molnar, Istvan

APPLICANT: Zirkle, Ross

APPLICANT: Cyr, Devon

APPLICANT: Goerlach, Joern

TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

FILE REFERENCE: 4-30582A

CURRENT APPLICATION NUMBER: US/09/335,409

CURRENT FILING DATE: 1999-06-17

NUMBER OF SEQ ID NOS: 30

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 22

LENGTH: 305

TYPE: PRT

ORGANISM: Sorangium cellulosum

US-09-335-409-22

Query Match 66.7%; Score 32; DB 3; Length 305;

Best Local Similarity 55.6%; Pred. No. 1.4e+02;

Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GPHEITVAL 9

|||||

Db 250 GPHEVLVVL 258

## RESULT 34

US-09-568-102-22

Sequence 22, Application US/09568102

Patent No. 6346404

## GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James

APPLICANT: Molnar, Istvan

APPLICANT: Zirkle, Ross

APPLICANT: Cyr, Devon

APPLICANT: Goerlach, Joern

TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

FILE REFERENCE: 4-30582A

CURRENT APPLICATION NUMBER: US/09/568,102

CURRENT FILING DATE: 2000-05-10

PRIOR APPLICATION NUMBER: 09/335,409

PRIOR FILING DATE: 1999-06-17

NUMBER OF SEQ ID NOS: 30

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 22

LENGTH: 305

TYPE: PRT

ORGANISM: Sorangium cellulosum

US-09-568-102-22

Query Match 66.7%; Score 32; DB 3; Length 305;

Best Local Similarity 55.6%; Pred. No. 1.4e+02;

Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GPHEITVAL 9

|||||

Db 250 GPHEVLVVL 258

## RESULT 35

US-09-567-969-22

Sequence 22, Application US/09567969

Patent No. 6355457

## GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/567,969  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-567-969-22

Query Match 66.7%; Score 32; DB 3; Length 305;  
Best Local Similarity 55.6%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
||||| : |  
Db 250 GPHEVLVWL 258

RESULT 36  
US-09-568-480-22  
; Sequence 22, Application US/09568480  
; Patent No. 6355458  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,480  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-480-22

Query Match 66.7%; Score 32; DB 3; Length 305;  
Best Local Similarity 55.6%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
||||| : |  
Db 250 GPHEVLVWL 258

RESULT 37  
US-09-568-486-22  
; Sequence 22, Application US/09568486  
; Patent No. 6355459  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,486  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-486-22

Query Match 66.7%; Score 32; DB 3; Length 305;  
Best Local Similarity 55.6%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
||||| : |  
Db 250 GPHEVLVWL 258

RESULT 38  
US-09-568-472-22  
; Sequence 22, Application US/09568472  
; Patent No. 6358719  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,472  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-472-22

Query Match 66.7%; Score 32; DB 3; Length 305;  
Best Local Similarity 55.6%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 GPHETITAL 9  
||||| : |  
Db 250 GPHEVLVWL 258

RESULT 39  
US-09-567-899-22  
; Sequence 22, Application US/09567899  
; Patent No. 6383787  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goerlach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/567,899  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17



```
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-899-22

Query Match      66.7%; Score 32; DB 3; Length 305;
Best Local Similarity 55.6%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      1 GPHETITAL 9
      ||||| : |
Db      250 GPHEVLVL 258

RESULT 40
US-09-328-352-4994
; Sequence 4994, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4994
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4994

Query Match      66.7%; Score 32; DB 4; Length 307;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GPHETIT 7
      ||||| : |
Db      124 GPHLTIT 130

RESULT 41
US-09-902-540-11852
; Sequence 11852, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 11852
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-11852

Query Match      66.7%; Score 32; DB 4; Length 315;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      2 PHETITAL 9
      ||||| : |
Db      35 PHETVEAV 42

; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-899-22

Query Match      66.7%; Score 32; DB 3; Length 305;
Best Local Similarity 55.6%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      1 GPHETITAL 9
      ||||| : |
Db      250 GPHEVLVL 258

RESULT 40
US-09-328-352-4994
; Sequence 4994, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4994
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4994

Query Match      66.7%; Score 32; DB 4; Length 307;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GPHETIT 7
      ||||| : |
Db      124 GPHLTIT 130

RESULT 41
US-09-902-540-11852
; Sequence 11852, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 11852
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-11852

Query Match      66.7%; Score 32; DB 4; Length 315;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      2 PHETITAL 9
      ||||| : |
Db      35 PHETVEAV 42
```

```
RESULT 42
US-09-489-039A-12482
; Sequence 12482, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709,2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12482
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12482

Query Match      66.7%; Score 32; DB 4; Length 383;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GPHETITAL 9
      ||||| : |
Db      276 GVHETITEL 284

RESULT 43
US-08-986-963-2
; Sequence 2, Application US/08986963
; Patent No. 5958730
; GENERAL INFORMATION:
; APPLICANT: Rosteck Jr., Paul R.
; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
; FILE REFERENCE: Ptsy
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: US
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/986,963
; APPLICATION NUMBER: US/08/986,963
; FILING DATE: December 8, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-11763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317/276-3334
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 425 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-986-963-2

Query Match      66.7%; Score 32; DB 2; Length 425;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      2 PHETITAL 9
      |||||
Db      340 PHETFLAL 347

RESULT 44
US-09-583-110-4561
; Sequence 4561, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 4561
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-4561

Query Match      66.7%; Score 32; DB 4; Length 429;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 PHETITAL 9
      |||||
Db      344 PHETFLAL 351

RESULT 45
US-09-007-484-2
; Sequence 2, Application US/09007484
; Patent No. 6072032
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: No. 6072032e1 FtsY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/007,484
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/007,484
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T.
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-007-484-2

Query Match      66.7%; Score 32; DB 3; Length 430;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 PHETITAL 9
      |||||
Db      345 PHETFLAL 352

RESULT 46
US-09-309-682-2
; Sequence 2, Application US/09309682
; Patent No. 6214348
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: No. 6214348e1 FtsY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/309,682
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/007,484
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T.
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-309-682-2

Query Match      66.7%; Score 32; DB 3; Length 430;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 PHETITAL 9
      |||||
Db      345 PHETFLAL 352

RESULT 47
US-09-107-433-2686
; Sequence 2686, Application US/09107433
```

```
TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-007-484-2
```

```
Query Match      66.7%; Score 32; DB 3; Length 430;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      2 PHETITAL 9
      |||||
Db      345 PHETFLAL 352
```

```
RESULT 46
US-09-309-682-2
; Sequence 2, Application US/09309682
; Patent No. 6214348
; GENERAL INFORMATION:
; APPLICANT: Black, Michael T.
; TITLE OF INVENTION: No. 6214348e1 FtsY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/309,682
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/007,484
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T.
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-309-682-2
```

```
Query Match      66.7%; Score 32; DB 3; Length 430;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      2 PHETITAL 9
      |||||
Db      345 PHETFLAL 352
```

```
RESULT 47
US-09-107-433-2686
; Sequence 2686, Application US/09107433
```

Patent No. 6800744  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTICS  
AND THERAPEUTICS  
NUMBER OF SEQUENCES: 5206  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: <Unknown>  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/ 085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 2686:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 430 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...430  
SEQUENCE DESCRIPTION: SEQ ID NO: 2686:  
US-09-107-433-2686  
Query Match 66.7%; Score 32; DB 4; Length 430;  
Best Local Similarity 75.0%; Pred. No. 2e+02; Mismatches 0; Indels 2; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0; Indels 2; Gaps 0;  
QY 2 PHETITAL 9  
DB 345 PHETITAL 352  
RESULT 48  
US-09-540-236-2280  
Sequence 2280, Application US/09540236  
Patent No. 6673910  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR  
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709, 2005-001  
CURRENT APPLICATION NUMBER: US/09/540,236  
CURRENT FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 3840  
SEQ ID NO 2280  
LENGTH: 431  
TYPE: PRT

ORGANISM: M.catarrhalis  
US-09-540-236-2280  
Query Match 66.7%; Score 32; DB 4; Length 431;  
Best Local Similarity 66.7%; Pred. No. 2e+02; Mismatches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GPHETITAL 9  
DB 179 GEHQAITAL 187  
RESULT 49  
US-09-413-814-72  
Sequence 72, Application US/09413814  
Patent No. 6225064  
GENERAL INFORMATION:  
APPLICANT: Geesellschaft fuer Biotechnologische Forschung mbH  
APPLICANT: Bristol-Myers Squibb, Co.  
APPLICANT: Beyer, Stefan  
APPLICANT: Bloecker, Helmut  
APPLICANT: Brandt, Petra  
APPLICANT: Cino, Paul M  
APPLICANT: Dougherty, Brian A  
APPLICANT: Goldberg, Steven L  
APPLICANT: Hofle, Gerhard  
APPLICANT: Mueller, Joachim  
APPLICANT: Reichenbach, Hans  
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
TITLE OF INVENTION: heteropolyketide compounds  
FILE REFERENCE: PCT/US 95/23535  
CURRENT APPLICATION NUMBER: US/09/413,814  
CURRENT FILING DATE: 1999-10-07  
EARLIER APPLICATION NUMBER: DE 198 46 493.2  
EARLIER FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 72  
LENGTH: 492  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-413-814-72  
Query Match 66.7%; Score 32; DB 3; Length 492;  
Best Local Similarity 55.6%; Pred. No. 2.4e+02; Mismatches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GPHETITAL 9  
DB 437 GPHEVLVVL 445  
RESULT 50  
US-09-328-352-7300  
Sequence 7300, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 7300  
LENGTH: 536  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-7300  
Query Match 66.7%; Score 32; DB 4; Length 536;  
Best Local Similarity 66.7%; Pred. No. 2.6e+02; Mismatches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPHETITAL 9  
| | : | | |  
Db 285 GERQAITAL 293

Search completed: July 27, 2005, 12:42:50  
Job time : 53 secs